final report

Project code: V.LDL.1902

Prepared by: Dr Alex Ball, Kimbal Curtis, Dr Paul Blazkiewicz and Johan Boshoff
Rural Analytics, Livestock Dynamics and University of New England

Date published: 29/11/2019

PUBLISHED BY
Meat and Livestock Australia Limited
Locked Bag 1961
NORTH SYDNEY NSW 2059

Developing aggregated insights for LDL

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

Executive summary

Livestock Data Link (LDL) was developed by Meat and Livestock Australia as a platform for accessing and benchmarking carcase data, grid information, animal health parameters and for reporting on non-compliance rates to market specifications for the sheep and cattle producers and processors in the Australian red meat industry. A key delivery outcome from LDL is the provision of carcase feedback to producers and processors that enables the prediction, understanding and manipulation of production and processing efficiency.

The Integrity Systems Company (ISC) commissioned a series of projects that were designed to provide recommendations and examples of how to improve the functionality and decision support capacity within LDL. One of those opportunities is to better translate carcase data collected at slaughter and carcase grading into benchmarked information that guides producers and processors to make more informed decisions on ways to improve carcase compliance to market specifications and pricing grids. The most effective way of communicating information to producers and processors is to create a series of insights that allow the end user to visualise carcase attributes by several different filters or aggregation that are based on fields that exist within LDL.

For this project, ISC identified objectives that would enable the ISC to make informed decisions on what insights would be the most useful for producers and processors given the structure of the LDL data sets, what formats those insights should be displayed in and finally what types of interactive dashboards could be created based on knowledge of the fields that are within LDL. The specific project objectives were: -

I. Develop a comprehensive list of aggregated data insights that might be considered as candidates;
II. Provide a shortlist of candidate insights selected following consultation with ISC, the LDL User Advisory Committee (UAC), key industry stakeholders and producers;
III. Provide example results for the shortlisted insights, along with appropriate documentation. Estimated value of the 'insights' in improving compliance.
IV. Provide a final report outlining the key activities undertaken in the project and the recommended set of aggregated data insights.

In August 2019, the project team received an extract from the LDL database of carcase data from four beef and two sheep processors. This data had been deidentified by ISC to ensure that processor confidentiality was maintained. The data supplied contained information on carcase traits collected either at slaughter or AUSMEAT grading for sheep and cattle, those data being hot standard carcase weight (HSCW) and a measure of fat depth (Fat-score for sheep and P8 for cattle). In addition, there was carcase data collected during MSA grading that included Rib-Fat, ossification, eye muscle area and the MSA index. Processors provided additional production characteristics including sex, dentition and feed type (grass vs grain) but those fields were often incomplete. There was no direct data available for breed.

This project contributed to the development of a survey that was sent to significant number stakeholders of LDL and the ISC. Unsurprisingly 71% of producer respondents to that survey indicated that obtaining insights on carcase compliance from LDL would be useful in making informed management decisions. More than two thirds of the producers indicated that they were prepared to share data and that the key insights that they required were geographical (location) and time-based benchmarks of carcase attributes. A number indicated that breed would be important, however this is currently not captured in the LDL database.

This project then used the deidentified data to provide a logic map for aggregation and insights of information by time, location, breed, sex and PIC number of the producer for both sheep and cattle. As one of the key requirements of the project was to systematically explore a range of different visualisation types, a range of options for visualisation of aggregated data into insights based on standard statistical terms and graphical representations were explored. Advantages and disadvantages for each of the graphic types were identified based on the interpretability of the graphic, its value to end user and the difficulty of visualisation.

The project identified that for single trait or factor insights, a combination of tables, histograms and line graphs is the optimum combination, particularly when benchmarking against time. For insights that had two or more factors and a range of production characteristics a standard combination chart of a scatterplot and a histogram would be most appropriate and best understood by livestock producers. This combination provided the end-user with the ability to visualise the variation, range and outliers that exist within their data relative to other producers across both time and geographical dimensions and by production characteristics.

Given that many of the traits in LDL have both a phenotypic and genetic correlation or relationship and often these relationships differ with production characteristics and time, the project team proposes the scoping and development of a dynamic dashboard which allows the end user to select various combinations of traits(factors), production variables and location options. Three sub-types of the dashboards have been sketched to illustrate different options for visualisation of the insights. Examples of these dashboards using different combinations of graphic options to visualise the insights are presented in the report.

Determining the value of the insight is problematic, given that all insights are based on retrospective analysis of historical consignments. A grid control chart was explored as an alternative graphic display for showing the value of an insight. It appears that grid control charts could be a useful addition to LDL.

Throughout the project, several recommendations for the ISC management team were identified. These are shown in the following table. These recommendations are not in rank or priority order.

Number	Recommendation
1	That a combination of a table and line chart or histogram be used for visualisation of one trait when aggregated against a production characteristic or time series
2	That a scatterplot / histogram combination be used as standard graphical interface for both variable and factor (aggregates) when combining two traits in LDL
3	That factors are drawn in a hierarchical order from the most abundant to the least abundant to enhance visualisation when overlying a time or location time series
4	That a dynamic dashboard comprising three sub-types is considered for implementation in LDL. This would enable an end-user to customise visualisation of data for one to multiple traits
5	That the dashboard concept be expanded to incorporate grid functionality by aligning thresholds or boundaries within the grids to the 3rd sub-type of dashboard
6	That consideration is given to the construction of a grid control chart when producers enter grids as a way of graphically demonstrating the value of an insight.
7	That LDL consider the ability of producers to upload production characteristics that may not be collected at slaughter or grading, to enhance LDL insights

Table of contents

1.1 Table of Figures

1.2 Table of Tables

2 Background

2.1 Aggregated insights for LDL

Within the Australia red meat industry, increasingly sophisticated value chains provide greater opportunities for the capture, aggregation and reporting of information to help producers and processors make more informed management and purchasing decisions respectively. Meat and Livestock Australia (MLA) via the Integrity Systems Company (ISC) have created and implemented the data reporting platform of Livestock Data Link (LDL). LDL provides the opportunity to translate data collected at slaughter and carcase grading into reports that guide producers and processors to make more informed decisions on ways to improve carcase compliance to market specifications and pricing grids.

To provide such functionality LDL needs to provide a comprehensive framework of information aggregation, analysis and utilisation within value chains. The framework must be responsive to differences in end-user expectations depending on the dynamics of the value chain including market specifications, brand options and resulting grid structures. To do this requires LDL to create opportunities for visualisation and interpretation of carcase attributes by several different filters or aggregation fields that exist in LDL. A further outcome of this project is the application of the aggregation framework (insight) and how this is expected to lift the compliance of carcases into higher value or higher compliance brackets.

3 Project objectives

The initial phase of this project was a systematic understanding of the existing capability of LDL including a review of current needs of end-users of LDL (via an on-line survey). This review and the survey feedback indicated that the following areas should be evaluated and explored: -

- Components of the benchmarking framework: Determination of the specificity of regionality, seasonality, production dynamics, customised group and value chain analysis.
- Benchmarking summary statistics and advanced analysis: Examination of the statistical dimensions of the traits within sectors identified above. This will include means, standard deviation, coefficient of variation, outlier analysis, correlations with other traits and where appropriate contour analysis.
- Visual representations (principally through statistical analysis including principal components and contour mapping) of the possibilities for change and modification within key production scenarios.
- Standardised visualisation: One of the key restrictions to improvement of compliance is that presentation of information in absolute or raw terms often is misleading in terms of the ability to understand the relative position of an individual animal, a consignment or an aggregated group relative to a defined population. It also leads to over or under emphasis on traits or derived traits due to scale rather than variance within a population. Investigate the potential of methodologies such as normalisation, standardisation for variance and coefficients of variation to rescale traits so that they can be more appropriately compared

3.1 Overall project objectives

The following are the stated project objectives: -

1. Provide a comprehensive list of aggregated data insights that might be considered as candidates;
2. Provide a shortlist of candidate insights selected following consultation with ISC, the LDL User Advisory Committee (UAC), key industry stakeholders and producers;
3. Provide example results for the shortlisted insights, along with appropriate documentation. Estimated value of the 'insights' in improving compliance.
4. Provide a final report outlining the key activities undertaken in the project and the recommended set of aggregated data insights.

4 Methodology

4.1 Data used to develop examples of aggregated insights

The project team was provided with a database of records from LDL by the ISC in August 2019. The data came from four beef and two sheep processors. The data sets were deidentified prior to being made available to the project team to avoid any potential biases and issues with producer and processor confidentiality. The database, as supplied, contained separate tables for cattle and sheep carcases that identified the carcase traits and derived variables such as percent lean meat yield (LMY%) and the MSA index, provided some indication of production type, dentition and sex; and finally provided dates for both slaughter (sheep and cattle) and grading (cattle). An overview of this data is presented in the following sections.

The data was loaded into a SQL database and examined with standard database and statistical software.

4.1.1 Rationale for LDL data aggregation (insights)

In the formative stages of the project, the team identified that there were several motivations for producers to engage with LDL. These motivations influenced the type of descriptive statistics required and then logic for the insights required. In response, the team has constructed a logic diagram of LDL interaction and a logic diagram for data insights for both sheepmeat and cattle. These are included in section 4.

4.1.1.1 Cattle

For cattle, there were carcases with kill dates from 6-Jan-2014 through to 25-Jul-2019.

The table below shows basic statistics for the primary measurements recorded for cattle carcases. Only 13% of the 2.198 million cattle carcases in the database had a full complement of these measurements. Some of the key features of the data received included: -

- Male cattle accounted for 57% of the 1.576 million carcases for which a sex was recorded.
- Over two thirds of the cattle were sourced from PICs located within NSW.

Table 1. Mean, standard deviation and 25th and 75th percentiles for measurements on cattle carcases (n=281,182)

Variable	Mean	Standard deviation	25th percentile	75th percentile
HSCW (kg)	328.7	55.3	291.5	366.0
Rib-fat (mm)	5.60	3.35	3.34	7.86
Fat depth P8 (mm)	16.3	7.2	11.9	21.1
Eye muscle area(cm)	76.5	12.3	68.2	84.9
AUS marbling (score)	1.56	1.12	0.80	2.31
MLA index (score)	58.2	4.4	55.2	61.1
LMY% (%)	61.1	2.6	59.4	62.9

Of those cattle with the measurements in the table above

- 28% were grain finished, as compared with the 72% finished on grass.
- 23% had received HGP

4.1.1.2 Sheep

For sheep, there were just under 13 million carcases with kill dates from 18-Dec-2013 to 24-Jul-2019. The following table gives summary statistics for the measurements reported for the ovine carcases.

- Figure 1 shows the distribution of fat scores.
- Sex was reported for less than 0.05% of the carcases.
- Dentition was report as ‘0’ for 92% of the carcases (lambs).
- Only 34% of the carcases had information to identify a state of origin.

Table 2. Mean, standard deviation and 25th and 75th percentiles for HSCW and LMY(n=5,513,742)

Variable	Mean	Standard Deviation	25th percentile	75th percentile
HSCW (kg)	22.94	4.38	19.99	25.89
Fat score	3.09	0.94	3	4
LMY% (%)	57.43	2.14	55.99	58.87

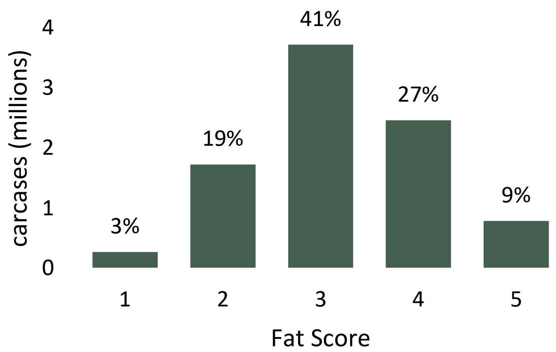

Figure 1. Distribution of lamb carcases between fat scores one through five. (n=8,976,632)

4.2 Defining the aggregated insights in LDL

Livestock Data Link is a data aggregation and reporting platform for the Australian Red Meat industries. It seeks to help producers and processors improve compliance rates to market specifications and thus reduce the costs of non-compliance. While uptake by industry is relatively low, the lack of information available to producers supplying to processors that are not contributing to LDL has been identified as a risk to success.

This project aims to identify and evaluate opportunities to provide candidate aggregated data insights that would be useful to producers, and where appropriate processors that utilise LDL.

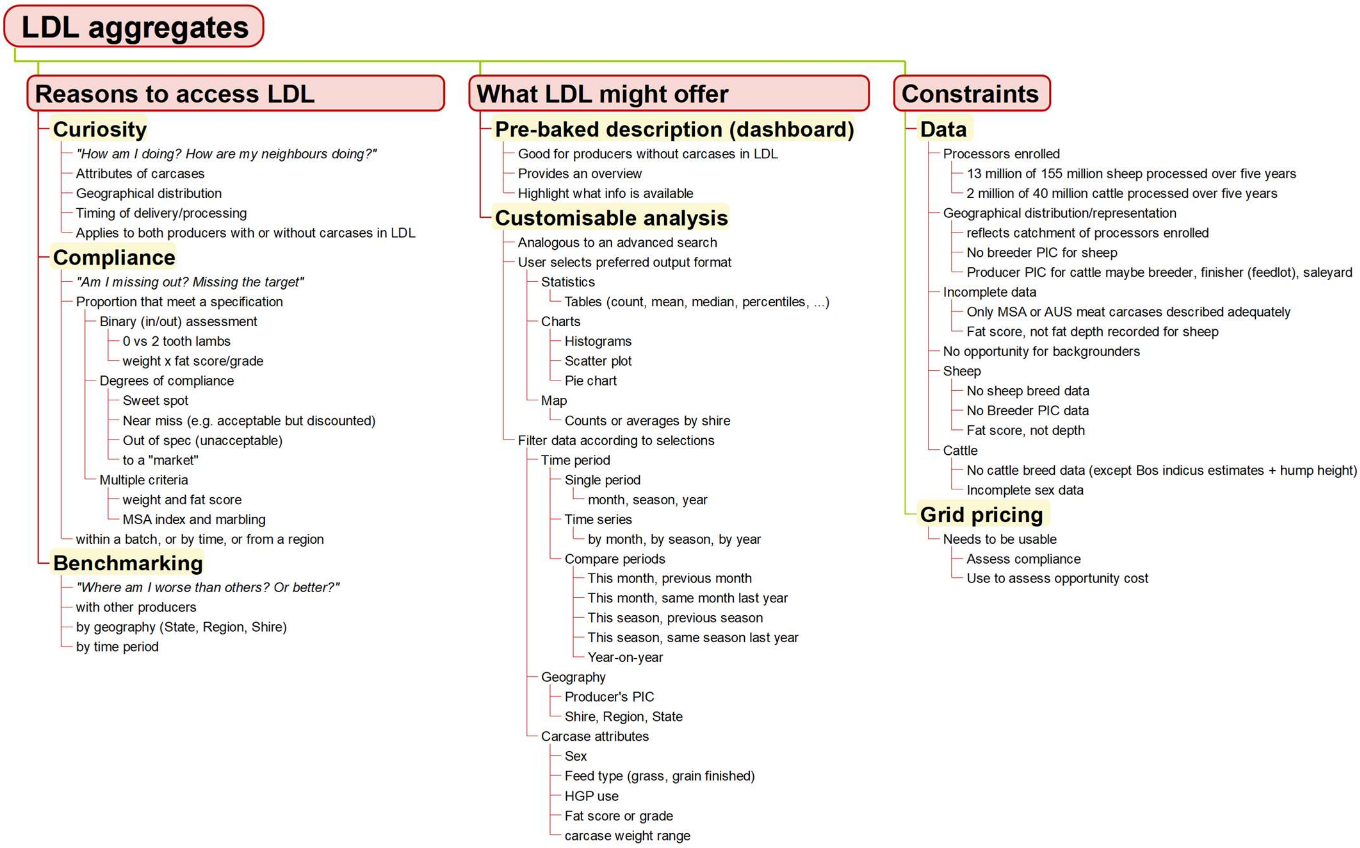

Figure 2. Understanding the logic behind the LDL aggregation project.

4.2.1 Logic for reasons that producers access LDL

It is hypothesised that there are three main reasons a producer (or processor) might access LDL.

- Curiosity: *How am I doing? How are my neighbours doing? Am I being competitive? What are others producing?*
- Compliance: *Am I missing out? What's my compliance rate?*
- Benchmarking: *Where am I worse than others? Or better?*

Naturally, these questions are most applicable to producers with carcase data in LDL, but they also apply to those who have supplied to other processors. While the latter group can't see the comparison on-screen, they can manually compare the feedback they have from their own processor with what LDL might provide.

4.2.1.1 Curiosity

Through LDL, producers could gain access to a broad spectrum of information on carcase attributes. Rather than just the actual measurements and appraisal of an individual's carcases, LDL can provide an industry context for those attributes identifying several statistics describing the typical values and their ranges. Data permitting, this information could be provided by geographical area and or by time period.

An advantage for a curious producer is that the information can be explored in private without social pressure from neighbours, friends and family.

4.2.1.2 Compliance

Assessing compliance is probably more useful to those with carcase data in LDL. They could apply the specification of their target market or markets to that data to determine and understand their level of compliance.

For the producer that doesn’t have data in LDL, they may still gain some benefit from observing the compliance rates of carcases from other producers in their region, though understanding which market a carcase was being delivered against is problematic.

Compliance includes both hit or miss criteria (e.g. lamb versus hogget, or exceeding or missing a threshold weight) as well as compliance against a scale where maximum benefit is received at a sweet spot (e.g. ideal carcase weight and fat depth combination, or MSA index and marbling) and value decreases with any departure from that combination.

4.2.1.3 Benchmarking

Benchmarking one producer’s production against his or her contemporaries can provide insights into what can be achieved and thus identify opportunities for improvement. This comparison can be made within a region, and for a period (month, season, year). While a direct comparison using data in LDL would be best, just having access to de-identified and aggregated data can form the basis for understanding where the opportunities for improvement might be found.

4.2.1.4 What LDL might offer

The following diagrams (Figure 3 and Figure 4) outline a process used to identify options for presentation of LDL data and insights. These are discussed in the following sections.

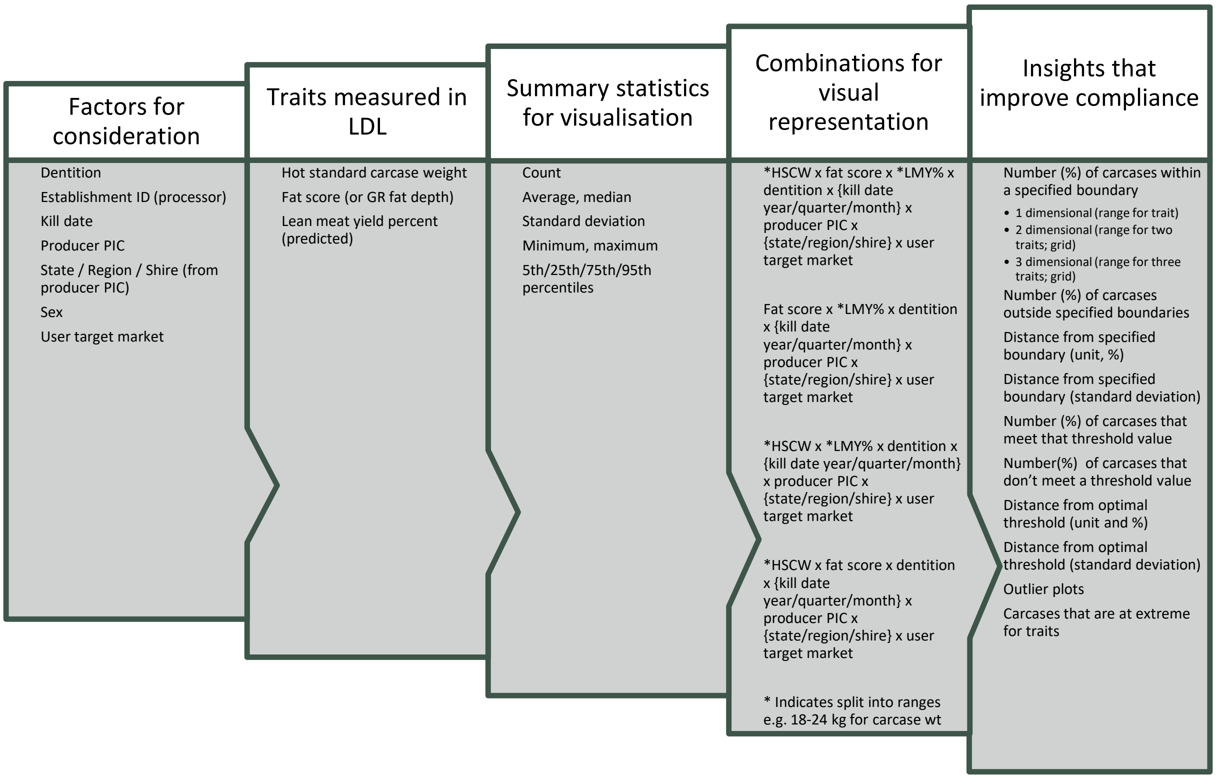

Figure 3. LDL Insight Map for sheepmeat.

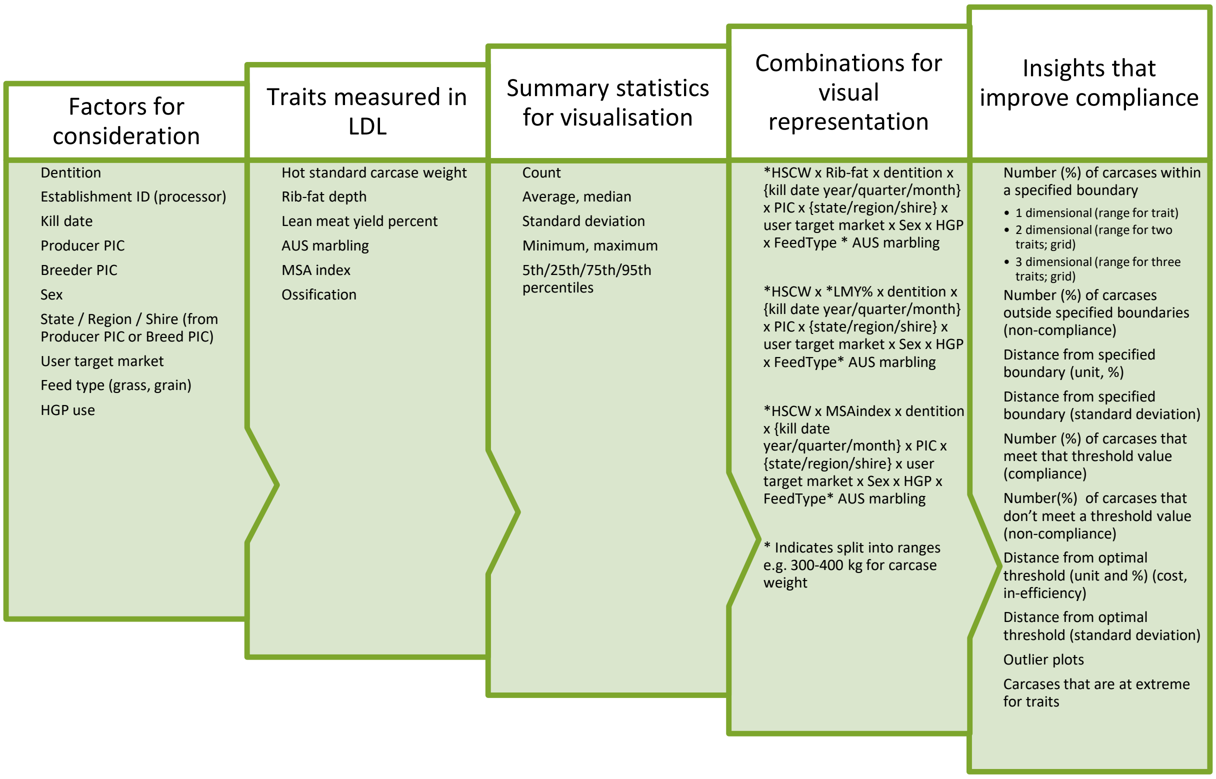

Figure 4. LDL Insight Map for cattle.

4.2.1.5 Pre-defined description (dashboard or snapshot)

A good starting point for a producer with little or no experience with LDL, might be the production of a summary of a range of relevant information. The starting point could be the user providing their PIC and indicating if their interest is in cattle or sheep.

The PIC would be used to identify if there is carcase data available for this producer. It would also be used to identify the target shire, region and state. This information could then be used to produce a snapshot of the carcase attributes recorded in LDL. The following tables suggest information that could be provided.

Table 3. Candidate information for a cattle dashboard

Attribute	Aggregates	Statistics used	Trait
Time period	User choice last 3 months, last 6 months, latest financial year, latest calendar year	Average or median, plus 10, 25, 75 and 90th percentiles and histograms or frequency distributions	Carcase weight Fat depth MSA Index Marbling score Predicted lean meat yield
Number of carcases processed	PIC*, shire, region, state.	Average or median, plus 10, 25, 75 and 90th percentiles and histograms or frequency distributions	Carcase weight Fat depth MSA Index Marbling score Predicted lean meat yield
Percent male (or percent female)	PIC*, shire, region, state.	Average or median, plus 10, 25, 75 and 90th percentiles and histograms or frequency distributions	Carcase weight Fat depth MSA Index Marbling score Predicted lean meat yield
HGP use Grass versus Grain Dentition	PIC*, shire, region, state.	Average or median, plus 10, 25, 75 and 90th percentiles and histograms or frequency distributions	Carcase weight Fat depth MSA Index Marbling score Predicted lean meat yield

Table 4. Candidate information for a sheep dashboard

Attribute	Aggregates	Statistics used	Trait
Time period.	User choice last 3 months, last 6 months, latest financial year, latest calendar year	Average or median, plus 10, 25, 75 and 90th percentiles and histograms or frequency distributions	Carcase weight Fat score Predicted lean meat yield
Number of carcases processed	PIC*, shire, region, state.	Average or median, plus 10, 25, 75 and 90th percentiles and histograms or frequency distributions	Carcase weight Fat score Predicted lean meat yield
Percent male (or percent female)	PIC*, shire, region, state.	Average or median, plus 10, 25, 75 and 90th percentiles and histograms or frequency distributions	Carcase weight Fat score Predicted lean meat yield
Dentition	PIC*, shire, region, state.	Average or median, plus 10, 25, 75 and 90th percentiles and histograms or frequency distributions	Carcase weight Fat score Predicted lean meat yield

4.2.1.6 Customisable analysis

This is where the real power of LDL can be expressed. It is somewhat analogous to an advanced search function.

The producer can choose a preferred output format (e.g. statistics, charts, maps), and then filter the data according to his or her interest.

- Statistics – these might include those proposed for the dashboard (count, mean/median, percentiles) but could be expanded (minimum, maximum, standard deviation or variance, measures of skewness).
- Charts – histograms or pie charts for single variables, scatter plots or scatter plots with histograms for showing the distribution and relationship between two variables (e.g. fat score/depth or predicted lean meat yield by carcase weight), line plots to show a trend over time.
- Maps – If enough carcases can be assigned to shires based on a producer or breeder PIC, then the geographical spread of supply could be mapped using dot density or bubbles to represent the source of livestock.

For the chosen format, the user would be presented with a set of input options. These would be used to filter the LDL data to a desired specification. These options might include:

- Time period – start time, finish time, interval (or step). These inputs would allow the user to choose a single period (month, season, year, year-to-date) or a time series.
- Region – PIC*, shire region or state
- Carcase attributes
  - Weight range
  - Fat depth range or fat scores (single or multiple scores)
  - Sex (not available for sheep)
  - Dentition (single or multiple levels)
  - Feed type (grass or grain, cattle only)
  - HGP use (cattle only)
  - MSA index (cattle only)
  - Marbling (cattle only)

- Hump height (cattle only)
- Ossification (cattle only)

Ideally, the user would be able to create more than one specification and be able to compare the results directly on screen.

4.2.2 Constraints

The information and insights that can be shared with LDL users is very much dependent on the data that is available. This is likely the largest constraint on the usefulness of LDL.

4.2.2.1 Processors enrolled

The LDL database provided to the project contains 2.2 million cattle carcases and 13 million sheep carcases from the five years from 2014-15 to 2018-19. These were reported by two sheep processors and four cattle processors. In the corresponding period, ABS report 155 million sheep and 40 million cattle were slaughtered in Australia. Expanding the level of coverage in LDL will increase the robustness of any analysis produced.

4.2.2.2 Geographical distribution or representation

With only a limited number of processors providing carcase data, the geographical distribution of those carcases caricatures their catchment rather than the livestock industry. As a result, some regions are significantly underrepresented, and thus data from those regions needs to be interpreted with caution.

In addition, breeder PIC was not available for sheep carcases limiting the usefulness of LDL to producers whose lambs pass through feedlots.

4.2.2.3 Other data

There is no sheep breed data and limited cattle breed information thus confounding benchmarking comparisons. It is not possible from the LDL data to know if a lamb comparison is between consignments of the same breed or across breeds.

For cattle there is some information for some carcases including estimated Bos indicus content and hump height. Beyond these indicators of tropical breed content, breed is unknown.

Fat score for sheep has some use, but fat depth would convey more information.

5 Results

5.1 Consultation with stakeholders

As stated in the objectives, one of the key outcomes of this project is to identify visual options for the presentation of aggregates of carcase data. This visualisation should be targeted towards users of LDL who are either livestock producers or processors. In order to determine what characteristics or levels of aggregation are required, a survey was sent to a large demographic of LDL and potential LDL users (see section 3.1.2).

At the close of the on-line survey (November 2019), there had been 154 respondents, however responses to each of the questions asked in the survey did vary. Several questions were asked within this survey that had direct relevance to this project. Approximately 68% of respondents indicated that basic insights of compliance within their region was either very useful or extremely useful. Less than 10% indicated that such information was of limited use or was not useful. Furthermore over 71% of producers indicated that data on health and carcase compliance would be useful in making better on farm decisions.

More than 60% of respondents were "willing" to share some information relating to their consignments of livestock. Notably 78% were willing to share breed type and 72% carcase, disease and defect compliance. A significant proportion (78%) of producers indicated that a comparison of compliance between consignments would be either very useful or extremely useful.

As shown in Figure 5, producers believed that there are key filters for aggregation that would improve comparisons between producers. These included comparing within region (to shire level), time period (week, month, season), breed, carcase attributes, disease and defect and processor. Notably 16% of the respondents wanted all the filters that were proposed.

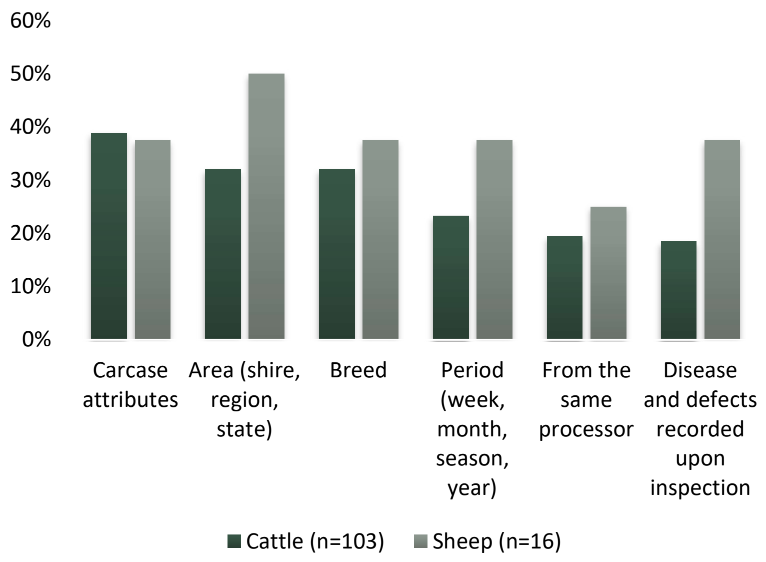

Figure 5. Survey responses for types of filters (aggregates) that producers would find useful in LDL.

The results of the survey have been used to determine initial options for aggregation both sheep and beef producers. These options are described within section 5.2.1.

5.2 Examples of aggregated industry insights for cattle and sheep producers

Data aggregation is a process in which information is gathered and expressed in a summarised form. The simplest way to display such summaries, would be to display statistics of interest in tabular form. The statistics of interest of the data may be the maximum, minimum, mean and standard deviation of the data set as well as many other types of statistics, an example is shown in Table 5.

Table 5. Aggregate statistics of hot standard carcase weight (HSCW), Rib-fat depth, predicted percent lean meat yield (LMY%) and MSA index for one producer comparing a single season with the previous season and the same season one year ago. In this table, Spring 2018 was used as the current season

		Current season (Spring 2018)	Previous season (Winter 2018)	Same season last year (Spring 2017)
HSCW	No. of carcases	259	65	50
	Mean (std. dev.)	318 (65)	277 (56)	318 (40)
	5th, 95th percentiles	189 – 446	164 – 390	236 – 399
Rib-fat depth	No. of carcases	259	65	50
	Mean (std. dev.)	5.3 (2.2)	4.5 (1.8)	4.9 (1.3)
	5th, 95th percentiles	0.9 – 9.6	0.5 – 8.4	2.1 – 7.7
Predicted LMY %	No. of carcases	259	65	50
	Mean (std. dev.)	61.2 (1.6)	61.7 (1.4)	61.6 (1.0)
	5th, 95th percentiles	57.8 – 64.5	58.6 – 64.7	59.4 – 63.7
MSA index	No. of carcases	259	65	50
	Mean (std. dev.)	52.0 (4.3)	58.4 (3.0)	52.6 (4.6)
	5th, 95th percentiles	43.2 – 60.8	52.0 – 64.9	43.0 – 62.1

When displaying aggregation statistics in a summary form, some information will always be lost. For example, the data distribution may be multi-modal, have skewness or kurtosis or other deviations from a single-mode normal-distribution. However, assessing whether information is lost cannot be completed without further investigation, or the use of additional statistics.

One solution to the problem of lost information is to graphically or visually display the data in different types of plots. Because visualisation can only occur within three spatial dimensions this makes it difficult to graphically represent data with more than 3 parameters, where each parameter has its own orthogonal spatial axis. There are techniques that try to get around this limitation. This report will examine how the multi-dimensional data in the LDL database can be grouped, summarised and displayed using various plotting techniques.

5.2.1 Aggregation or Filter fields (levels)

As indicated earlier in this report the LDL database provides the opportunity to collate a large range and volume of carcase data. Within that database there are several fields that, if collected by the processor, would enable significant insights into factors that affect compliance to both grid and market specifications. The purpose of such insights is to help producers and processors improve the quality of the carcases produced, reduce non-compliance by producing fewer carcases outside specification (thus improving production efficiency).

A key objective is to ensure that the output delivered by LDL is readily understandable and useful to users. To do this the visual representations or graphics need to be informative, interpretable and not overly complex.

In the review of the LDL database and supplied data tables, the project team identified the following four categories of data that are recorded in LDL.

i. **Time** – date of kill allowing aggregation over days, months, seasons, years or any other arbitrary time period.
ii. **Geography or location** – both breeder PIC and producer PIC are recorded for most cattle carcases and producer PIC for sheep carcases.
    a. Breeder PIC is the property where the animal was born and initially tagged with a NLIS tag. The same PIC may have sold/delivered direct to the processor.
    b. Producer PIC. Alternatively, the animal may have passed through one or more other PICs before being delivered to the processor.
    c. Each of these PIC options can be associated with a shire, region and state for comparison with other sources.
iii. **Factors** – attributes of the carcase that are not actual measures but rather they have a discrete set of possible values. Typically, they are not continuous variables and hence it is not sensible to calculate averages and variances. These include
    a. Sex – male, female
    b. Dentition – 0, 2, 4 tooth etc…
    c. HGP use (Yes or No)
    d. Feed type (Grass, Grain)
    e. Breed (including predicted Bos indicus content)
    f. Hump height
iv. **Variables** – these are values recorded across a range of values with a level of resolution dependent on the method of measurement (scale of the ruler, decimal places on a weighing device, etc.). Standard statistical analysis can be applied to these measures. They can be converted to factors (discrete values) by classifying them into ranges e.g. carcase weights over 400 kg. Variables include
    a. Hot standard carcase weight (hscw)
    b. Rib-fat depth for cattle and GR fat score for sheep and lambs
    c. Eye muscle area
    d. Lean meat yield %
    e. MSA index

For this project, it was assumed that the filters or aggregate levels to be investigated were either time based, geographically based that linked producer PIC to location or factors that impacted on production systems and therefore the resulting carcase characteristics. The variables included standard carcase attributes collected at processing for either AUSMEAT or MSA grading.

It was assumed that producers will be the main audience for insight data presented by LDL if achieving increased compliance to market specifications was the objective. The producer audience can be further separated into those who have carcase data that has been uploaded into LDL from participating processors and those whose animals have gone to other processors. Aggregate data could be useful to both groups with the former group having the added advantage of being able to benchmark their carcases against those of other producers.

5.2.2 Types of visualisation investigated

It is expected that the analysis and reporting will be made available through the LDL web portal. This information will largely consist of (a) raw data and appropriate statistical analyses, and (b) graphical presentation in a form that helps the user visualise the results. The following sections present analyses of the data provided and an exploration of graphical options with examples.

5.2.3 Graphical presentation of variables

Typically plot types are segmented into 2D and 3D plots, although they are always constructed on a two-dimensional surface such as a piece of paper or a computer screen. Common charts are the histogram, bar chart, pie chart, line chart, control chart, scatter plot, bubble chart, radar plot, star plot, parallel coordinate plot, contour plot, heat map, box plot and the violin plot. This list of plots is not exhaustive, and there are also specialist plots which are combinations of these various plots to display special characteristics of the data. For example, one can align and display multiple plots in a matrix or set of panels which are adjacent to each other.

Many of the prospective audience for the LDL information are practical hands-on producers or farm managers for whom tables of numbers and statistical analyses are not appealing and often do not convey direct interpretation of results. For these users, and in fact for many users, a graphical presentation with some guidance on interpretation is likely to be more readily appreciated and accepted. As a result, the project team has explored a range of options and noted the strengths and weaknesses of each. The following sections show some example plots using various techniques and discuss the advantages and disadvantages of each method displayed. The examples that follow use cattle data (hot carcase weight and MSA index) to illustrate options and note each format's strengths and weaknesses.

5.2.3.1 Line Chart

Line charts are two dimensional plots that are commonly used to display the evolution of a certain statistical parameter versus time. In this example, a query to calculate the max, min and mean of the sheep hscw values and the number of carcases, for each quarter for establishment id 7 (processor) was run. Figure 6, shows a two-panel plot of line charts, where the larger top panel shows the evolution of the max, min and mean hscw values, and the smaller bottom panel shows the number of carcases for each quarter.

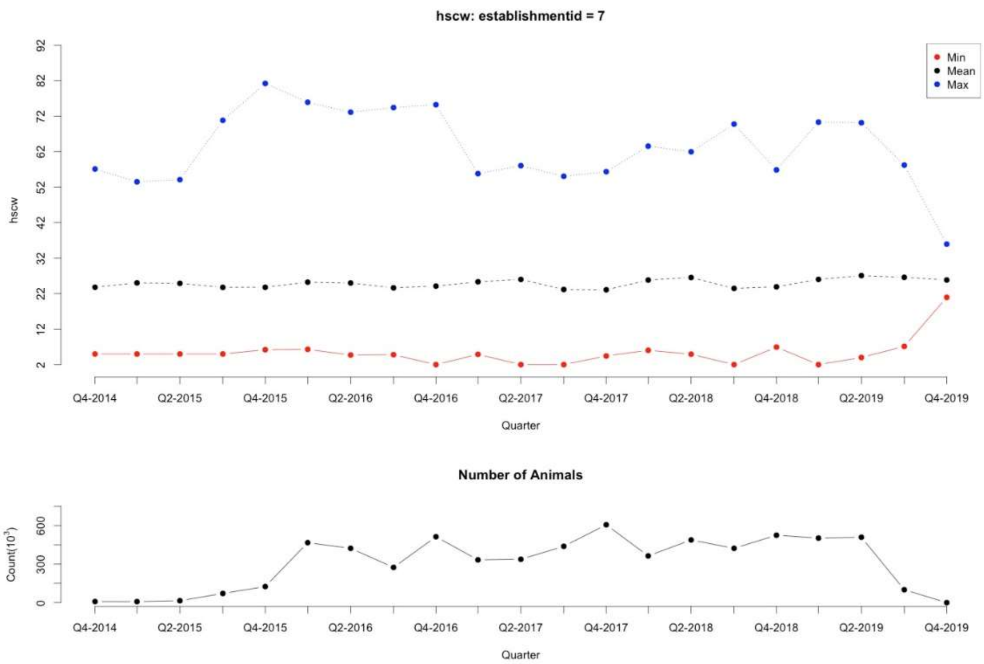

Figure 6. Hot standard carcase weight (hscw) quarterly time series for establishment ID 7.

In this example, using a two-dimensional format can effectively display 4-parameters, as three of the parameters have similar magnitudes and the same units of measure. Graphically displaying the data this way, allows for a quick visual inspection of the data and possibly identification of any features or problems.

For example, that wide range in the data is too high to be biologically reasonable. The raw data shows that the maximum value is 81.2 kg, the mean value is 24.8 kg and the minimum value is 2 kg. Assuming a dressing percentage of around 48% one can estimate the max, min and average weight of the sheep before slaughter. The calculated values would imply that the maximum sheep weight would be around 169 kg, the minimum would be 4.1 kg and the average weight would be 51.6 kg. These maximum and minimum values would seem to be outside the acceptance range of the abattoir, and possibly indicate a data entry error that should be investigated.

Table 6. Advantages and disadvantages of a line chart

Advantages	Disadvantages
• **Simplest presentation format** • **Almost universal understanding** • **Can be used for time series** • **Can present more than one series**	• Does not show distribution or outliers • Different scales can be confusing • Doesn't show correlation between factors

For several decades, the manufacturing industry has used "statistical process control" to improve product quality. Sources of product variation are typically minimised until the production process becomes stable. However, all sources of noise or variation cannot be fully removed. To monitor the process, typically a control chart is introduced, which is a line chart indicating the current unit status and the upper and lower boundaries of acceptance. If one of the units has a variation large enough to make it fall outside of the acceptable range, then an alert is produced, and the defective product is removed from the production line for inspection.

A similar thing can be done within LDL, where it could have a multi panel plot within a control chart indicating the max and min and mean values together with the upper and lower acceptance limits. That control chart could have two additional line-charts, one indicating the percentage of carcases breaching the upper limit and the second chart indicating the percentage of carcases breaching the lower acceptance limit.

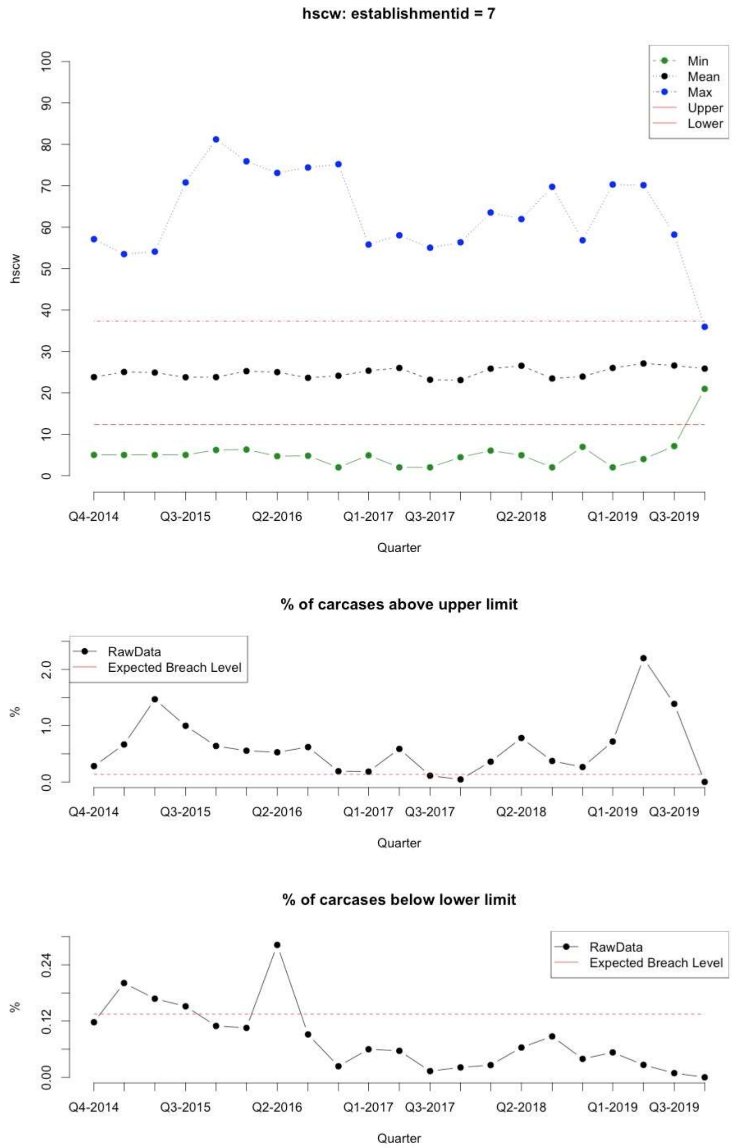

Figure 7. Example of a hscw control chart for establishment ID 7.

In Figure 7, the upper and lower control limits were set at 3 standard deviations from the mean hscw of the whole dataset. Setting such upper and lower limits, should lead to breaches of each limit around 0.135% of the time, if the hscw follows a normal distribution. In Figure 7, although the max and min values seem to be significant outliers, the percentage of the number of carcases breaching the lower limit hovered close to the expected level of breaches and then dropped below the expected level after Q3-2016. However, the upper limit was breached almost all the time in the 0.1 to 2% range, much more than would be expected for a normal distribution.

5.2.3.2 Histogram

A histogram is a representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable. To construct a histogram, the first step is to divide the entire range of data values into a series of intervals called bins or buckets, and then a count of how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable.

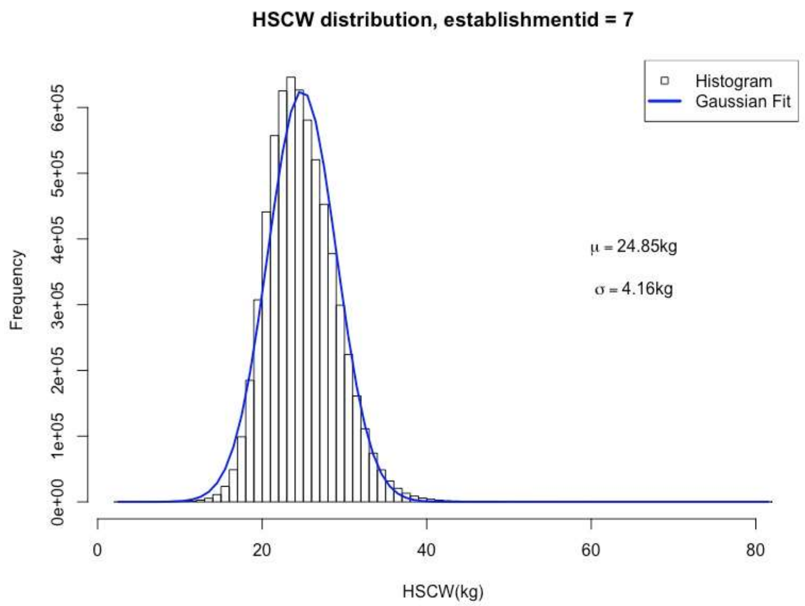

Figure 8. Hscw histogram for establishment id 7 LDL data.

In Figure 8, a histogram of the hscw data associated with establishment id 7 is shown. In addition, a line plot of a normal distribution is overlayed using the same mean and standard deviation of the hscw data. Here it is immediately obvious that the histogram only slightly deviates from a normal distribution.

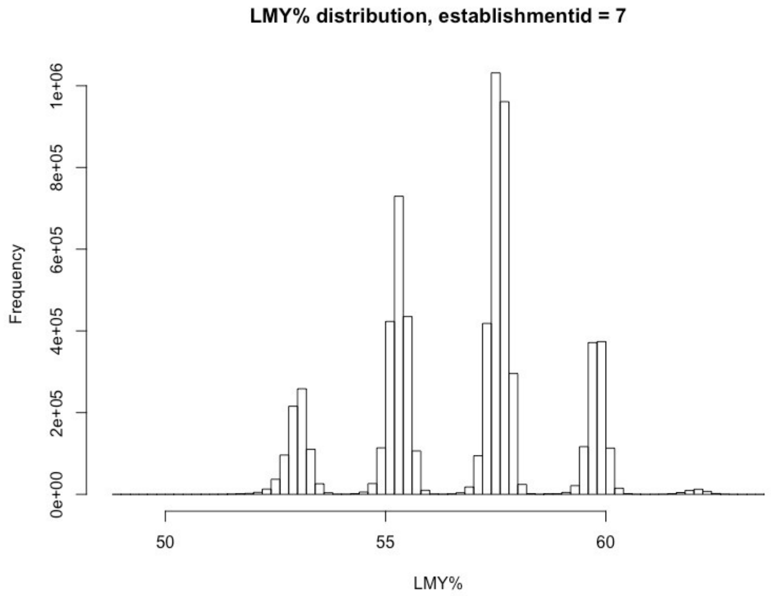

Figure 9. Predicted lean meat yield percentage histogram for establishment ID 7.

Figure 9 shows a histogram of the establishment ID 7 LMY% values. Intuitively one would have expected something resembling a normal distribution. However, one can see that the lean meat yield histogram has produced a multi-modal distribution where there is a peak associated with each fat score entered into the LMY% equation used in the LDL database. Producing a histogram like this would immediately warrant further investigate of the performance of the LMY% prediction algorithm. A repeat of this LMY% histogram using cattle data, is shown below in Figure 10.

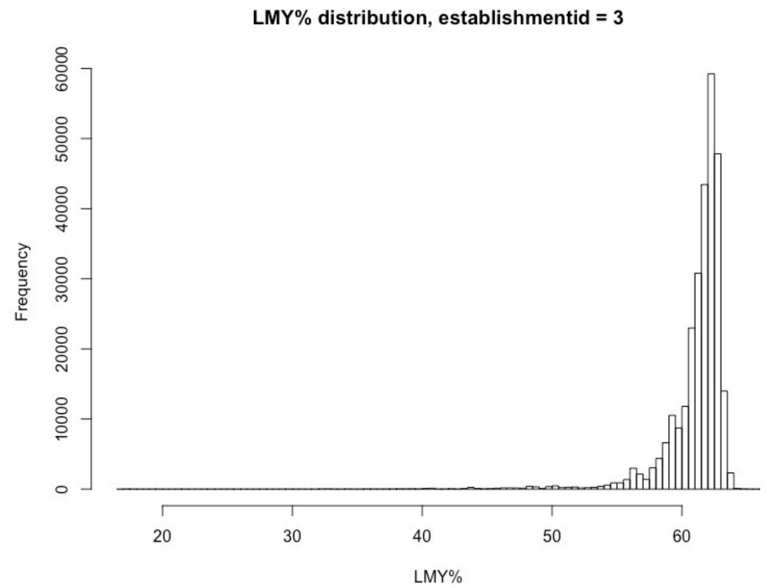

Figure 10. Predicted lean meat yield percentage histogram for establishment ID 3.

Again, one would have expected a normal distribution covering the range. But instead a negatively skewed distribution was obtained which caused unrealistically low values of LMY%. This again suggests that there is a problem with the cattle LMY% calculation model and methodology.

In addition, to producing single histograms on single parameters from single establishment IDs, one can make a set of panels where histograms are aligned along the dimension of interest to compare results. One example of this is shown in Figure 11, where four histograms are produced vertically with their horizontal axes aligned. The histograms are the hscw distribution of a producer compared against the total shire, region and state histograms. Also, the mean, standard deviation and number of animals are labelled on the side of each histogram for additional information.

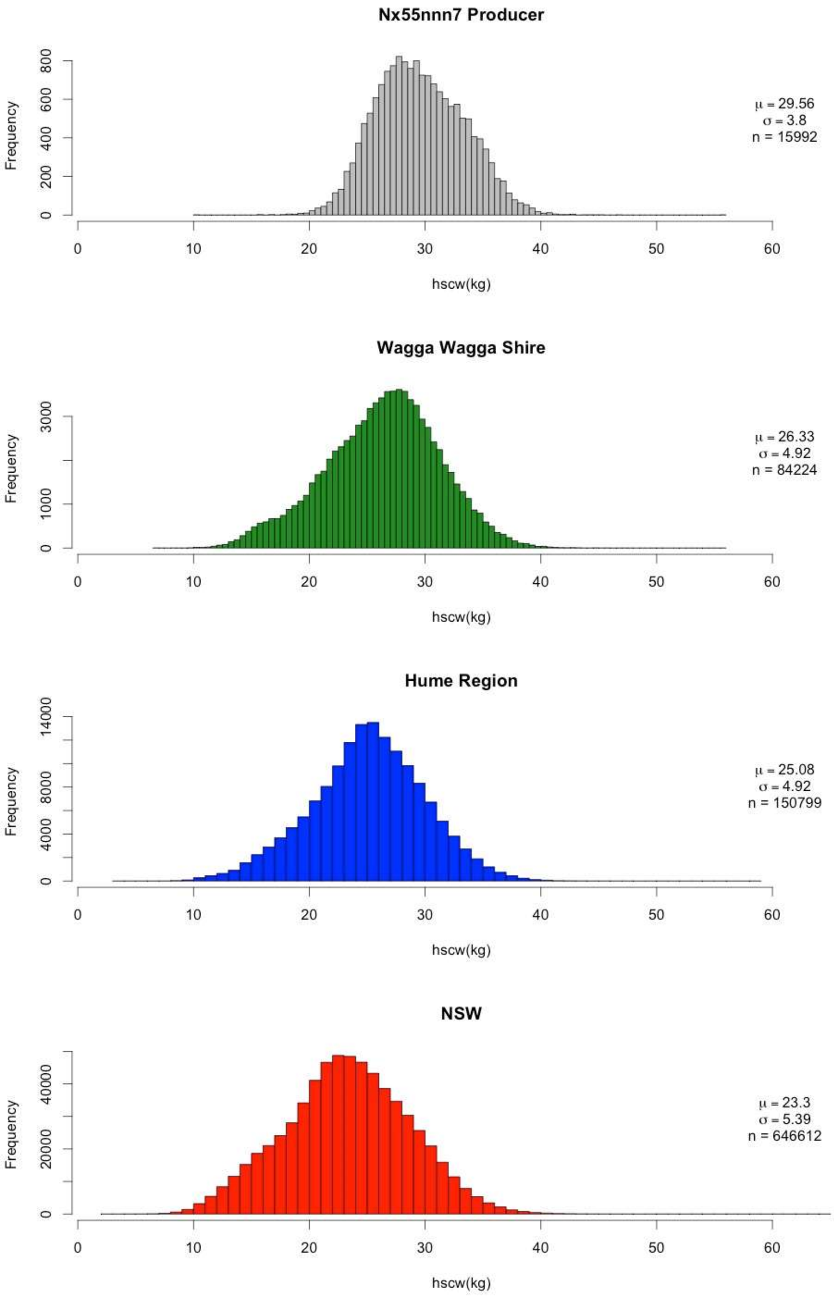

Figure 11. Histogram of hscw for individual sheep producer, shire, region and state.

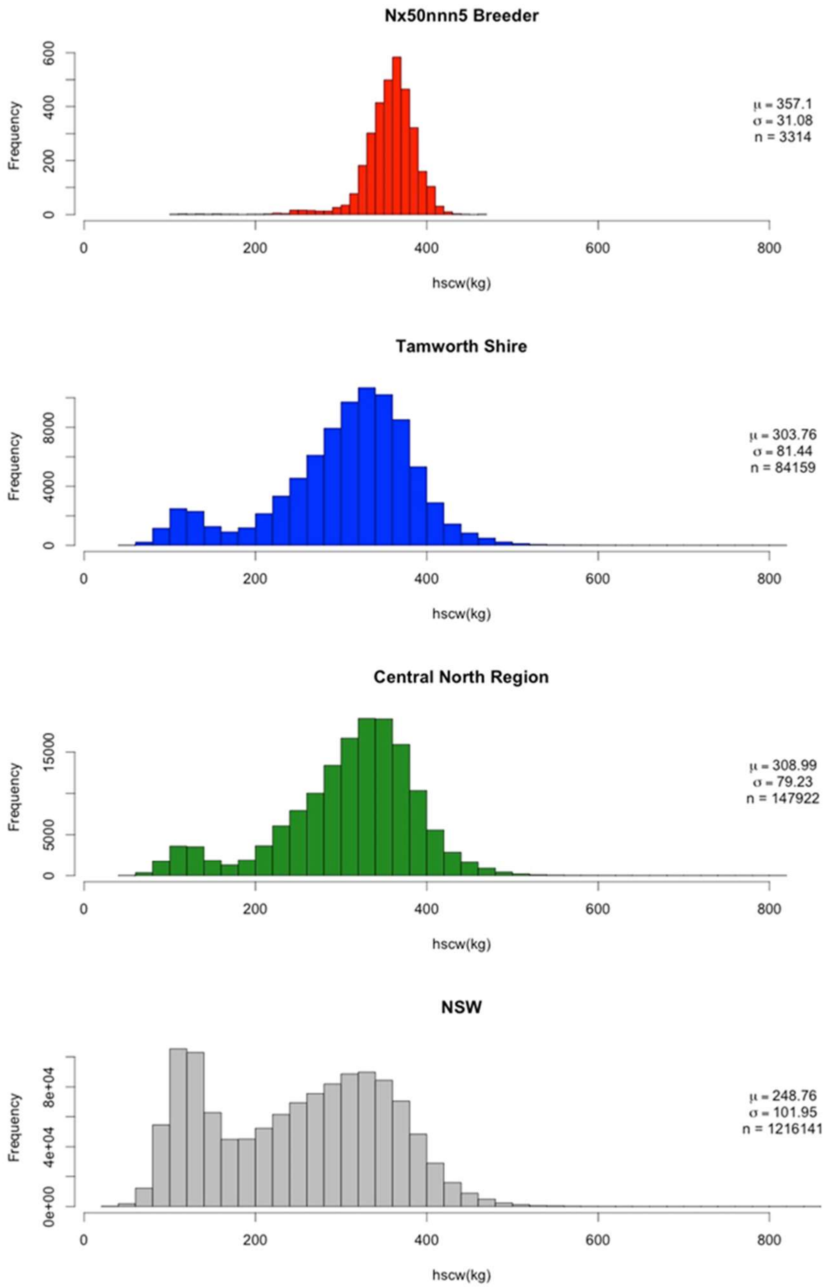

Figure 12. Histogram of hscw for individual cattle producer, shire, region and state.

Repeating this plot for a cattle producer gave very different results. For example, in Figure 12, the histogram of the individual producer is close to a single-moded normal distribution, possibly indicating that the producer focuses on one target market. However, looking down at the lower panels shows that when the region of interest is expanded to the total shire, region and state a more complicated multimodal distribution is observed. This is probably due to different animal types, production environment and end markets that all the animals are destined for. In such situations, it might be useful to allow the LDL user to further filter the animals to compare their own performance against others targeting the same or similar end markets.

Table 7. Advantages and disadvantages of a histogram

Advantages	Disadvantages
• **Simple format for presentations** • **Almost universal understanding** • **Shows distribution** • **Can present more than one series**	• Can't be used for time series • Different scales can be confusing • Multiple plots can hide data

5.2.3.3 Box Plot

Histograms are not the only method of displaying details of the statistical distribution of a parameter of interest. A common simpler method which loses some information is to use box plots. The simplest possible box plot displays the full range of variation (from min to max), the likely range of variation (the inter-quartile range; IQR), and a typical value (the median). It is not uncommon that real datasets display surprisingly high maximums or surprisingly low minimums, such occurrences are termed outliers. As shown in Figure 13, the box plot is a way of displaying five statistics about the distribution, that is the minimum, the first quartile, the median, the third quartile and the maximum value.

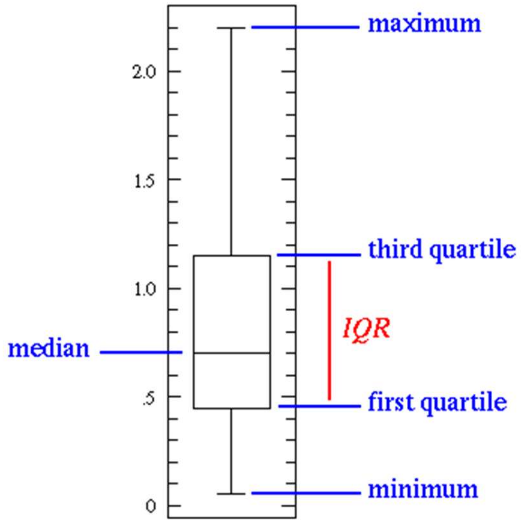

Figure 13. Details of a boxplot.

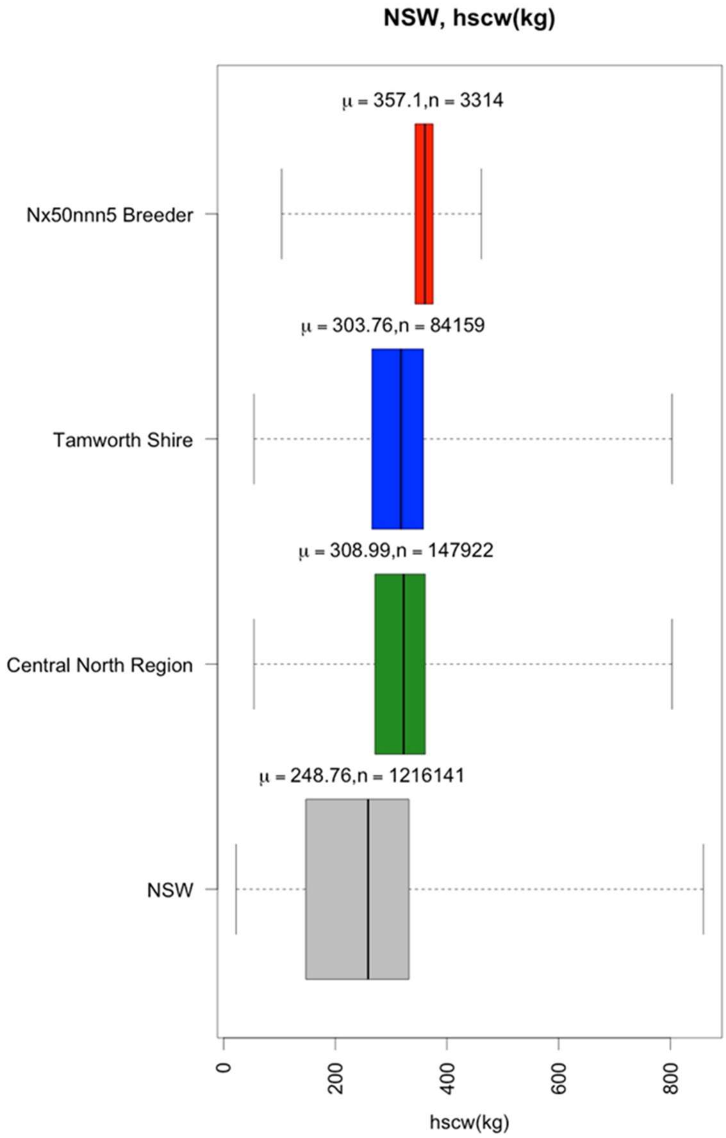

Figure 14. Horizontal boxplot of hscw for individual producer, shire, region and state.

Figure 14 uses the same hscw cattle data as in Figure 12, however it displays the results as a set of boxplots. The box plot shows that the state level inter-quartile range is much wider than the other categories but there is no knowledge or indication of the shape of the distribution which in Figure 12, was multi-modal.

Table 8. Advantages and disadvantages of a boxplot

Advantages	Disadvantages
• **Provides more extensive statistical information** • **Shows distribution** • **Can be used to show outliers** • **Can present more than one series**	• Not end user friendly • Different scales can be confusing • Requires some statistical calculations to be performed • Does not show shape of distribution

5.2.3.4 Violin Plot

A violin plot is another method of plotting the distribution of numeric data, which does not lose as much information as a boxplot. It is similar to a box plot, with the addition of a rotated kernel density plot on each side.

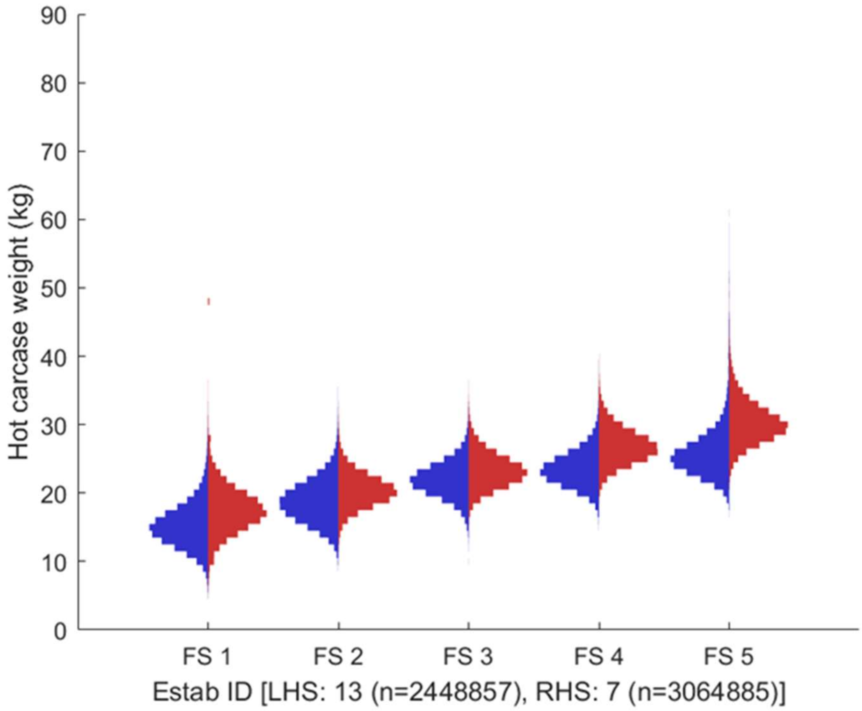

Figure 15. Violin plot of sheep hscw versus fat score for two establishment id's.

In Figure 15, is a more complicated violin plot, where the blue areas on the left side are the rotated density function for establishment ID 13. The red areas on the right side are the rotated density function for establishment ID 7, and there are five violin plots one for each fat score.

Table 9. Advantages and disadvantages of a violin plot

Advantages	Disadvantages
• **Provides a visual comparison of two distributions without one obscuring the other** • **A series of 'violins' can be used to present more than one series**	• Not in common use so may need extra support with interpretation • Only useful for a single measured variable • Second trait can only have two values • More difficult to produce

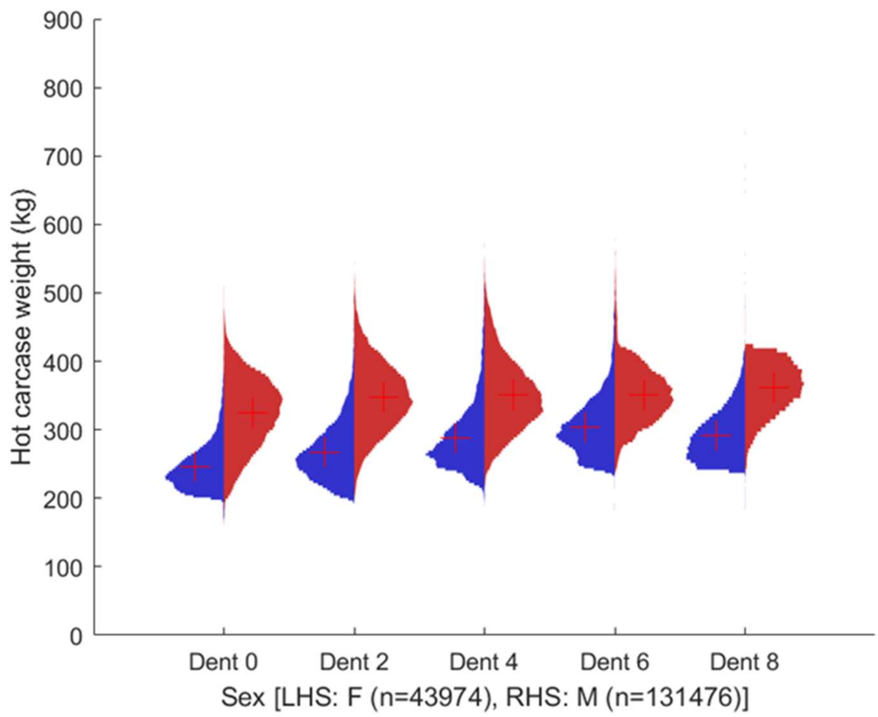

Figure 16. Violin plot of cattle hscw versus dentition for each sex.

In Figure 16, another example of a violin plot, the blue areas on the left side are the rotated density function for female carcases. The red areas on the right side are the rotated density function for castrate carcases, and there are five violin plots one for each dentition category. The red plus sign (+) indicates the median for each combination.

5.2.3.5 Scatter plot

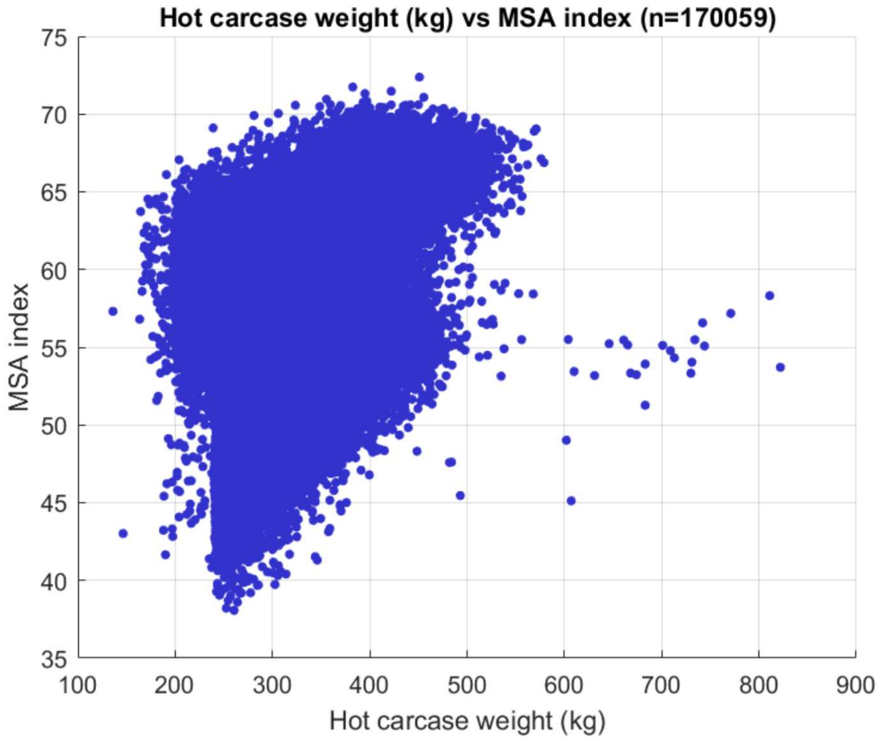

Figure 17. Example Scatterplot of MSA index vs hscw for cattle.

Perhaps the simplest way of presenting data that compares two measured variables is the scatter plot. Each point represents a single carcase in the database, in this example showing MSA index versus hot carcase weight. Several data series can be plotted on the same surface and distinguished by colour or symbol.

Table 10. Advantages and disadvantages of a scatter plot

Advantages	Disadvantages
• **Simple presentation** • **Almost universal understanding** • **Indicates any correlation between the variables** • **Can present more than one series**	• Hides density of points when n(number) is large • If a new series is added, it may hide points from a previous series • Different scales can be confusing

5.2.3.6 Bubble plot

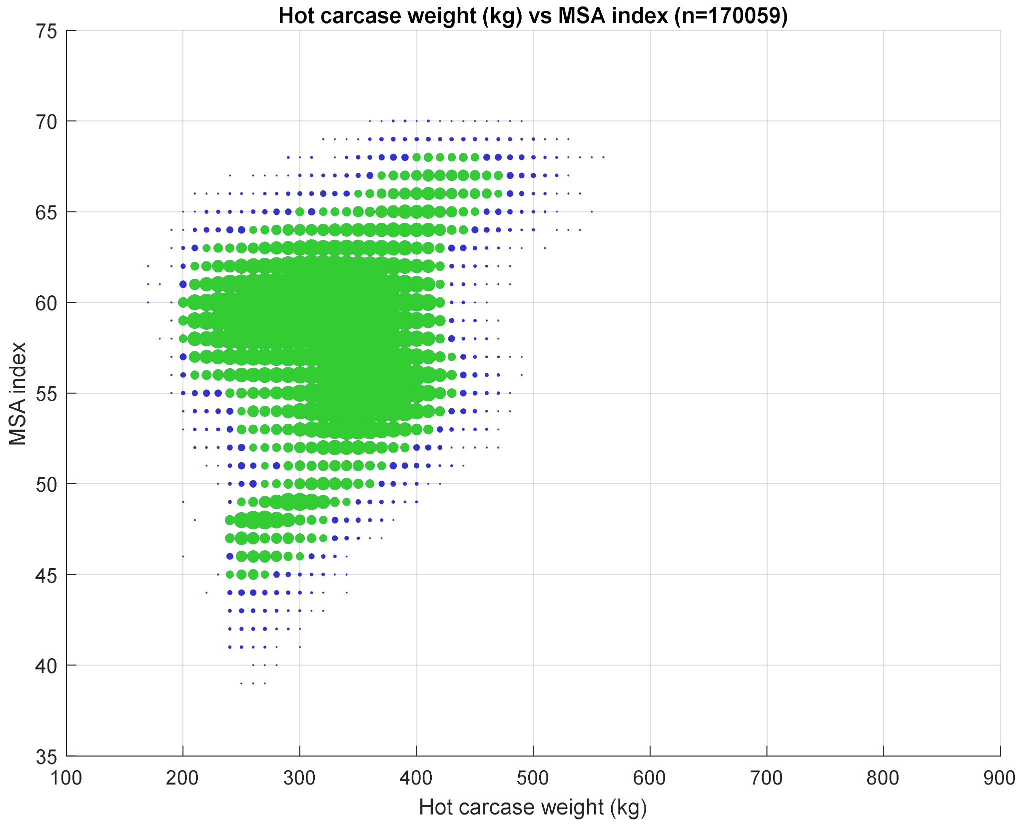

Figure 18. Example bubble plot.

A bubble plot has the same functionality as a scatter plot, but the relative size of the dots indicates the density of points (number of carcases). The bubble plot is complex to generate because the data must be binned into a matrix of values. This example presents the same data as in the scatter plot example. It shows where the points are most dense. Bubble plots have greater utility if there is a wide variation in density.

As a small variation on a standard bubble plot, this example has green dots representing the area accounting for 95% of all carcases and blue for the other 5%.

Table 11. Advantages and disadvantages of a bubble plot

Advantages	Disadvantages
• **Relatively simple presentation** • **Shows density of points when n is large** • **Relatively easy to interpret** • **Indicates any correlation between the variables** • **Useful if variation in density is important**	• Outliers are less obvious • Data must be binned • No advantage over scatter plot with small datasets • Not suitable for multiple series

5.2.3.7 Scatter plot with side histograms

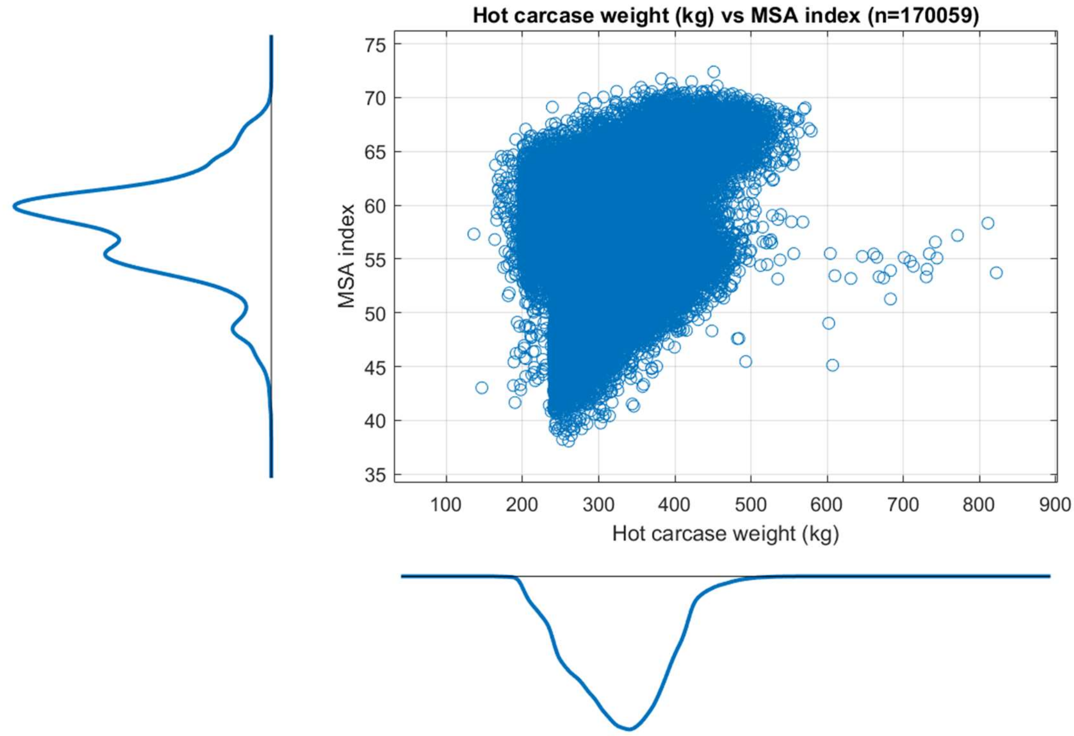

Figure 19. Example scatter-histogram plot

The above figure is a composite plot that combines the functionality of a scatter plot with side histograms (frequency distributions). It has the advantages of a scatter plot (shows the distribution of points across two variables) but also indicates for both variables the distribution of those variables across their ranges. Additional series can be added and distinguished by colour and or symbol.

Table 12. Scatter-histogram plot

Advantages	Disadvantages
• **All the benefits of a scatter plot** • **Adds a frequency distribution for each of the variables** • **Multiple series can be added**	• A little more complex to interpret • Requires more screen real estate, or the scatter plot must be smaller • Side distributions may not be useful when there are few data points

5.2.3.8 Contour plot

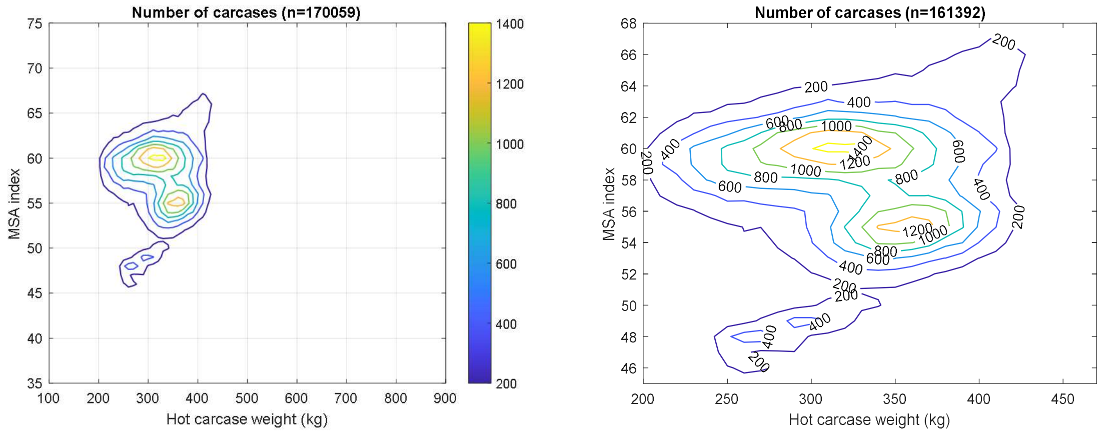

Figure 20. Contour plots. Left plot shows full extent of the data and includes a legend. Right plot zoomed in to limits of 95% of the points and lines are labelled with number of carcases.

Contour maps show the density of points on a two-variable surface. The lines represent points on the MSA index by carcase weight surface with the same number of carcases. In the left-hand chart, the lines are coloured according to the legend shown beside the plot. In the right-hand chart the lines are labelled. Unless the lowest contour is drawn at n=1, the outer contour does not show the full extent of the data.

Table 13. Advantages and disadvantages of contour plots

Advantages	Disadvantages
• **Simple presentation** • **Indicates both the spread of values (X by Y) and the density**	• Not suitable for low numbers of points • Data needs to be binned for some software • Can only display one series • Does not show outliers • Producers need to understand 2-dimensional data

5.2.3.9 Heat map

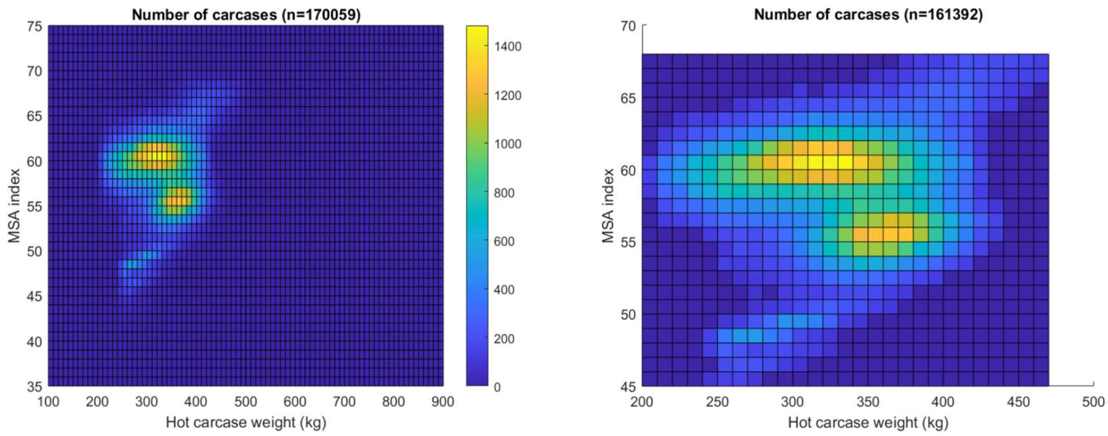

Figure 21. Heat maps versions of the contour plots in the previous figure.

Heat maps provide a graphical (and colourful; visual) image of the data but lack resolution to clearly show the relationship between the variables being plotted. The changing colours across the surface represent the density of points (carcases) at each combination of the X and Y variable.

Table 14. Advantages and disadvantages of heat maps

Advantages	Disadvantages
• **Useful to identify combinations of the X and Y variables where carcases are concentrated** • **Can identify hot spots or sweet spots easily**	• Not suitable where total numbers are limited • Not as effective at showing any correlation between variables • Data needs to be binned for some software • Outliers are not obvious • Needs to be shown in colour to get full appreciation of the data

5.2.3.10 3-D plot

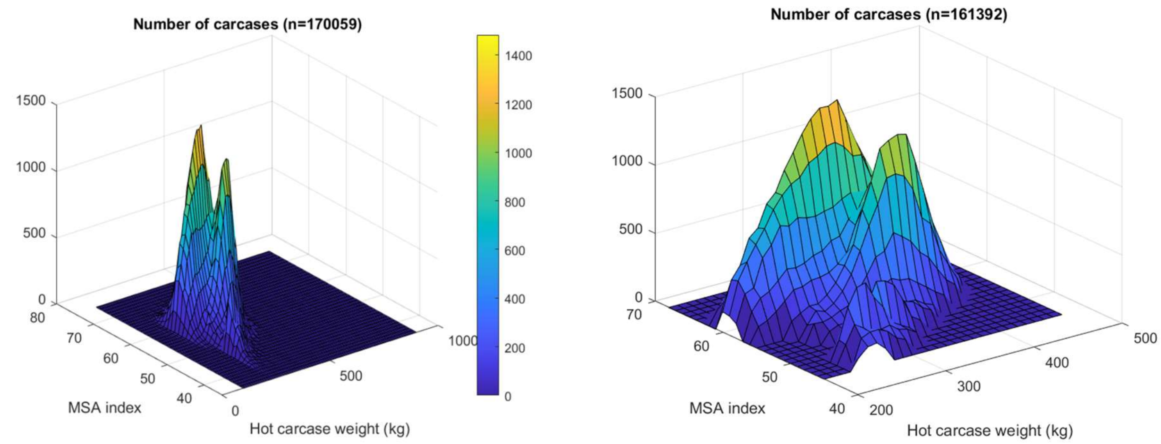

Figure 22. Surface plots. A false 3-D representation of the distribution of carcases on a two-variable matrix.

The surface or 3-D plots combine the contour and heat map representation into a virtual 3-D image. The apparent height above the X-Y plane represents the density of points. This kind of plot is more complex to interpret and for best effect, the user needs to be able to rotate the plot to view it from different directions. As a result, this format is more suitable in an interactive format.

Table 15. Advantages and disadvantages of surface (or 3-D) plots

Advantages	Disadvantages
• **Provides a visual representation of the data by both the variables on the X-Y plane and their density (height above the plane** • **Can identify hot spots**	• Some people struggle to visualise the false 3-D image • Difficult to identify the X-Y values for individual points on the surface • For best effect, the image needs to be interactive so the user can rotate it and zoom in/out • Only useful for larger datasets

5.2.4 Graphical representation of combination of variables and factors (aggregation)

From the preceding discussion, the project team identified that scatter plots and scatter plots with side histograms are likely the most useful and informative choice – scatter plots for their simplicity when comparing the range of two variables and any correlation between them, and scatter – histograms for the added information conveyed about the distribution of values for each variable.

Scatter plots and the other two-dimensional formats described in the preceding section have the additional advantage that they are like the pricing grids most commonly used by processors and so familiar to most producers. These grids show prices or premiums/discounts on a matrix, usually carcase weight versus a measure of fat.

Both options can convey additional utility when the points in the scatter plot and the frequency distributions are coded (different colours, symbols, line specifications) by other factors. As an example, the following charts (Figure 23) show the distribution of carcases as per the previous charts, but separately for cattle finished on grass and those finished on grain. These can be combined to show the overlap and the areas of difference (Figure 24). In this example, overlaying (aggregating) the data for grain and grass makes the comparison easier. The differences are particularly evident in the side histograms that show the grass finished cattle were on average lighter and had a higher MSA index. The MSA index distribution for the grain finished carcases shows a second lower broad peak with an MSA index well above the main peak of grass finished carcases.

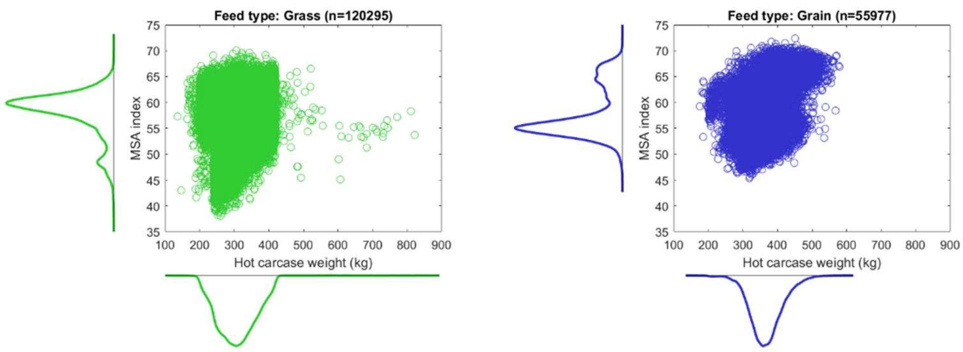

Figure 23. Scatter – histogram for MSA index versus carcase weight for grass finished (left) and grain finished (right) carcases.

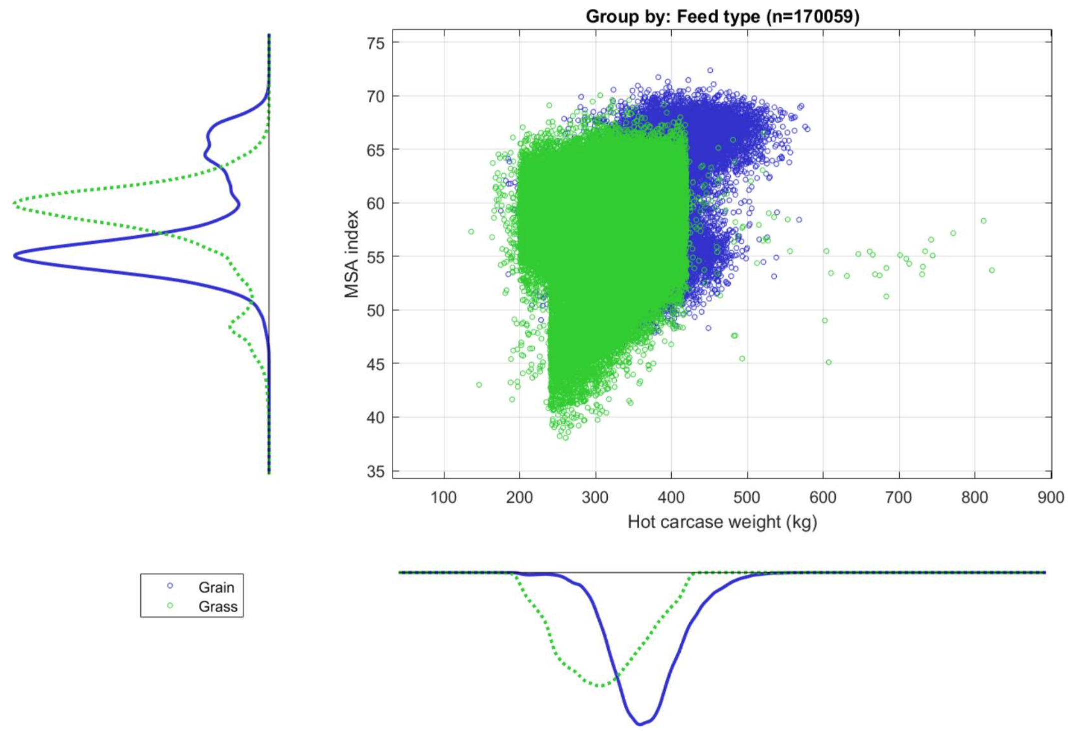

Figure 24. Scatter – histogram of MSA index versus carcase weight for both grass and grain finished animals.

Scatter – histogram plots can be used to explore the relationship between the two variables of the scatter plot and other factors (including feed type, HGP use, sex, dentition, AUS marbling, geography). The following figures present examples of these.

The decision of whether to draw separate plots for each level of a factor needs to be based on whether a second or subsequent layer covers an earlier layer. In the example, in Figure 24, the “grass” points hide much of the distribution of the underlying “grain” points, however some idea of their distribution can be deduced from the associated histograms. In this example, separate plots (Figure 23) might be more informative. In general, it is recommended that the factors be drawn from most abundant to least abundant to avoid one layer completely hiding others.

Figure 25 provides another example where 2nd, and 3rd layers hide the distributions of earlier series – again the side bar histograms help understand the distribution of each layer. But in this case the first layer mostly extends beyond the margins of the other layers.

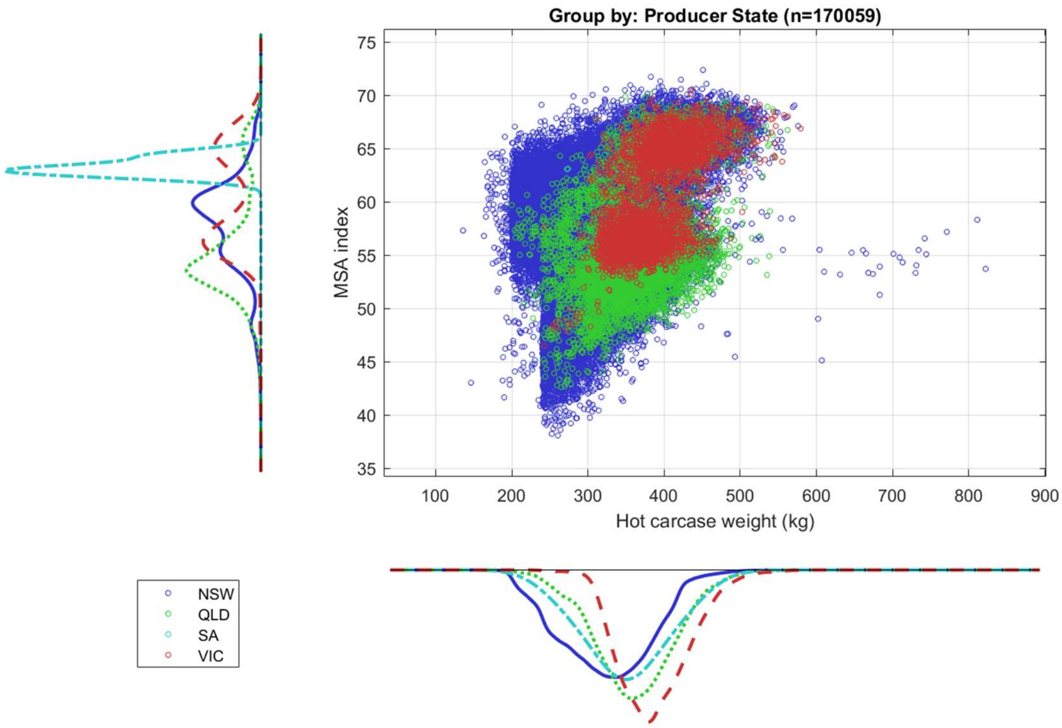

Figure 25. Distribution of carcases on an MSA index by carcase weight surface and coded by State of the producer delivering the animal to the processor.

With data from a geographical hierarchy (State, Region, Shire, PIC), it is crucial that layers are drawn in order from largest area to smallest because all the data in a smaller area (e.g. Shire) must, by definition, be in an enclosing area (e.g. State or Region). An example of this is shown in Figure 26 where carcases from a specific PIC (drawn in aqua) are overlayed on all carcases from the Shire (Moss Vale, green), the region (Cumberland, red) and the State (NSW, dark blue).

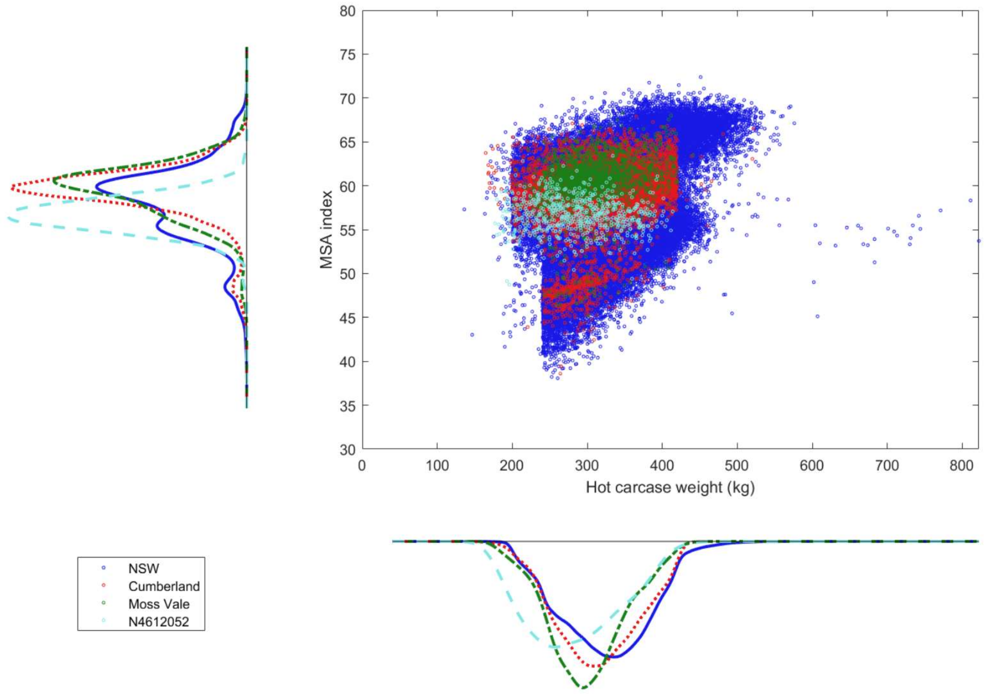

Figure 26. Distribution of carcases on an MSA index by carcase weight surface and coded by producer PIC[1], Shire, Region and State.

The number of levels (values) of a factor needs to be considered. If there are too many levels, then the picture can become so fuzzy that it is difficult to interpret. An example is shown in Figure 27, where the breeders of the animals are spread over seven different states.

[1] The PIC has been de-identified, and any resemblance of the value displayed to a real PIC is co-incidental.

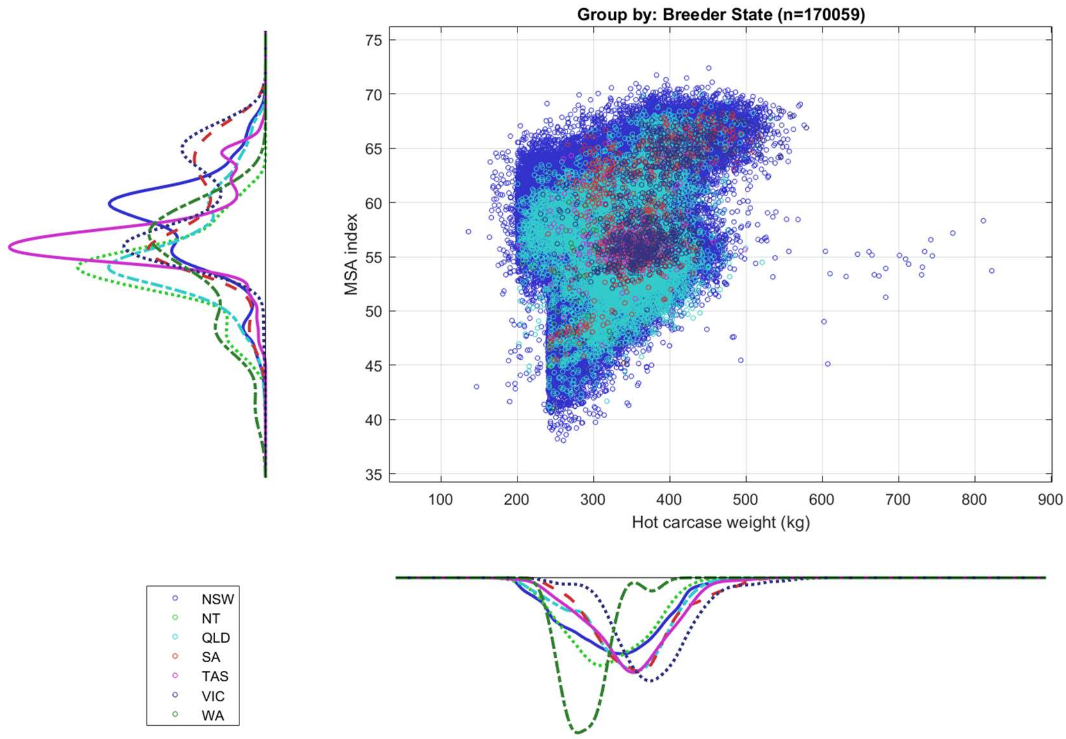

Figure 27. Distribution of carcases on an MSA index by carcase weight surface and coded by the State of the breeder of each animal.

5.2.5 User interactions

Where complex plots are displayed, the user may benefit from being able to simplify what is shown. For example, rather than showing all levels of a factor, the user might benefit from selecting a subset that is of interest.

Figure 28 shows the distribution of cattle carcases on a percent lean meat yield (LMY%) by hot standard carcase weight (HSCW) surface. The carcases were colour coded by the reported dentition. With six possible values reported for dentition, overlap of layers makes it difficult to interpret the relationships.

A simplification of that example is shown in Figure 29 where only carcases of two dentition levels are shown. In this example, dentition '0' and dentition '6' are compared. While the LMY% distributions are very similar, carcases with the higher dentition level ('6') were generally heavier than those with a dentition level of '0'. A trivial example, but one that illustrates the need for the user to have the option of simplifying the chart.

In offering the option to modify the levels displayed, it is recommended that plots be rendered with the same scales for the axes and that colours / symbols for each series (e.g. dentition level) are consistent no matter what combination is displayed.

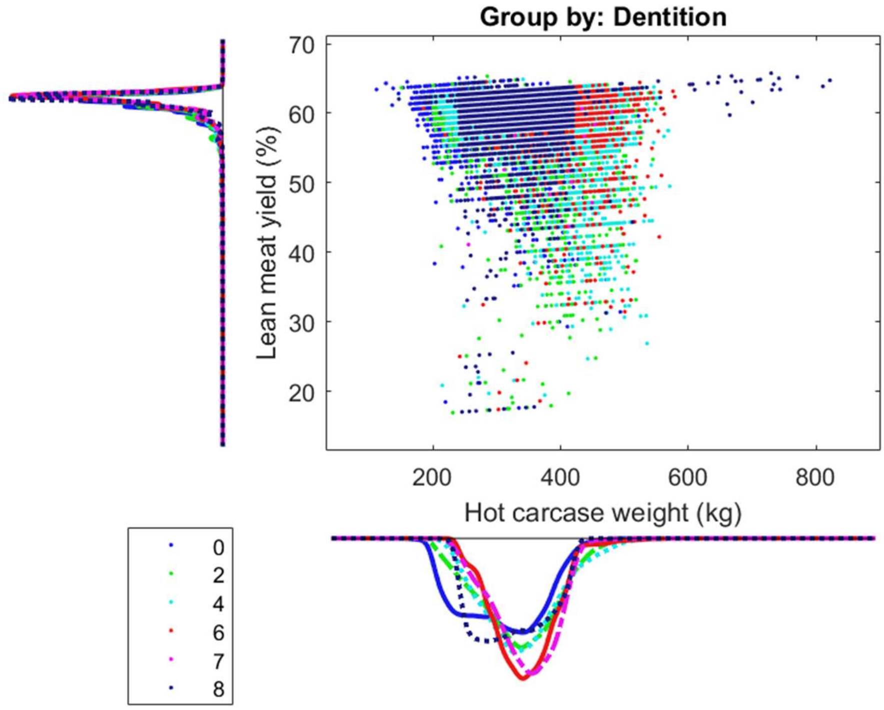

Figure 28. Distribution of carcases on a percent lean meat yield by carcase weight surface and colour coded by dentition.

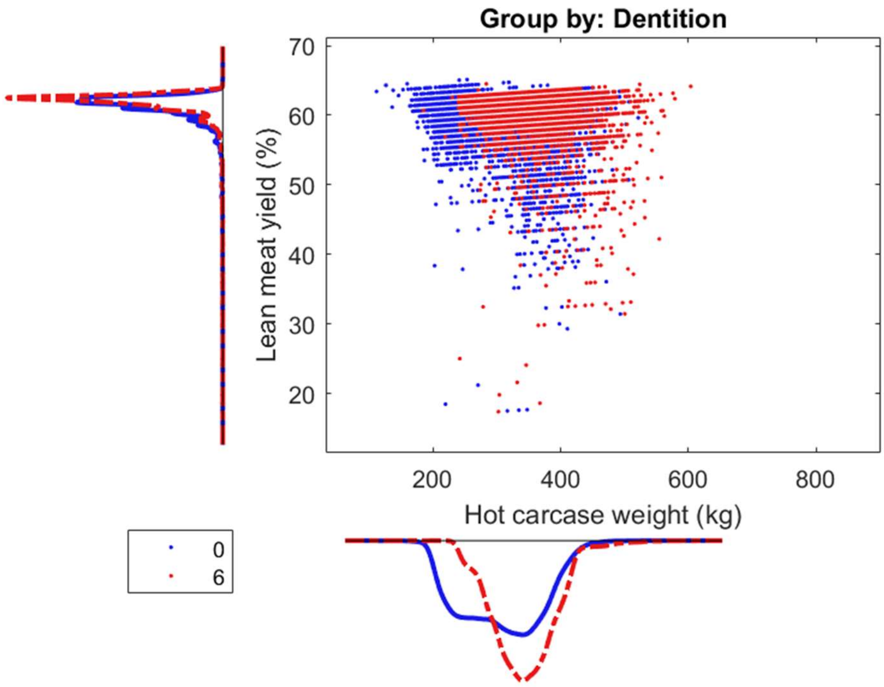

Figure 29. Distribution of carcases on a percent lean meat yield by carcase weight surface and colour coded by dentition, and restricted to only two levels, '0' and '6'.

5.2.6 Geographical hierarchy

The provision of PIC for the producer supplying the animal to the processor and the PIC for the property where the animal was bred are used to identify the carcases produced by a producer user of LDL. In addition, the PIC can be used to identify carcases from the same shire, region or state.

In demonstrating this geographical hierarchy, the following figures use a cluster of four charts showing the relationship between hot carcase weight (on the X axis) and (a) lean meat yield, (b) rib-fat cold, (c) eye muscle area and (d) MSA index.

The first figure shows all the data for carcases sourced from producers in NSW (Figure 30). The following three figures build on the NSW data with overlays for the Central North region (Figure 31), for the Tamworth shire (Figure 32) and for a PIC[2] within the Tamworth shire (Figure 33).

This sequence benchmarks the carcases from a single PIC with those from the same shire, region and state. This sequence could be further refined to only show carcases for a specified period (e.g. this year, last year). The first three of these figures, State, Region and Shire, might be provided to a producer who is yet to have any carcases recorded in LDL. The shire displayed could be chosen by the user or based on the user's PIC.

In addition, if having all four geographical levels included in the plot makes it too complex, layers can be hidden as in Figure 34.

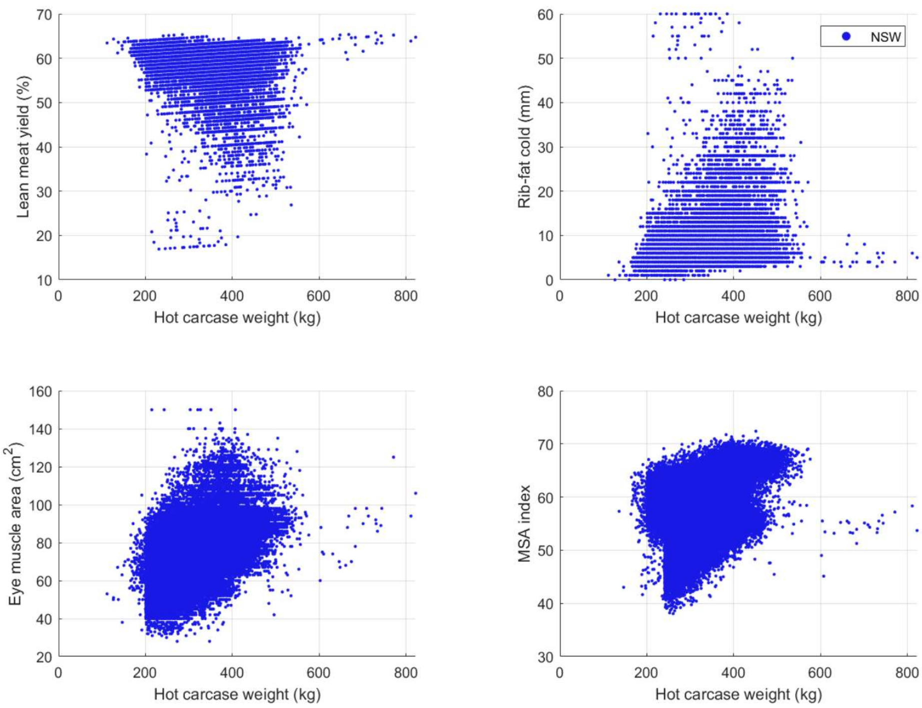

Figure 30. Lean meat yield (top-left), rib-fat cold (top-right), eye muscle area (lower-left) and lean meat yield (lower-right) versus hot carcase weight for carcases sourced from producers in NSW.

[2] The PIC has been de-identified, and any resemblance of the value displayed to a real PIC is co-incidental.

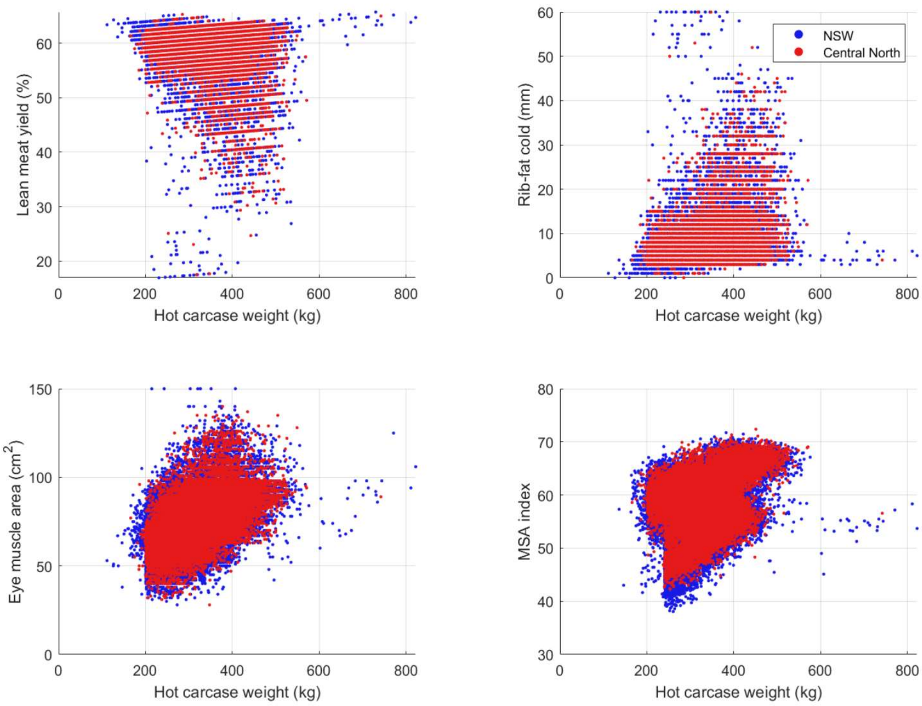

Figure 31. As for the previous figure, but with carcases from the Central North region overlaid on the NSW points.

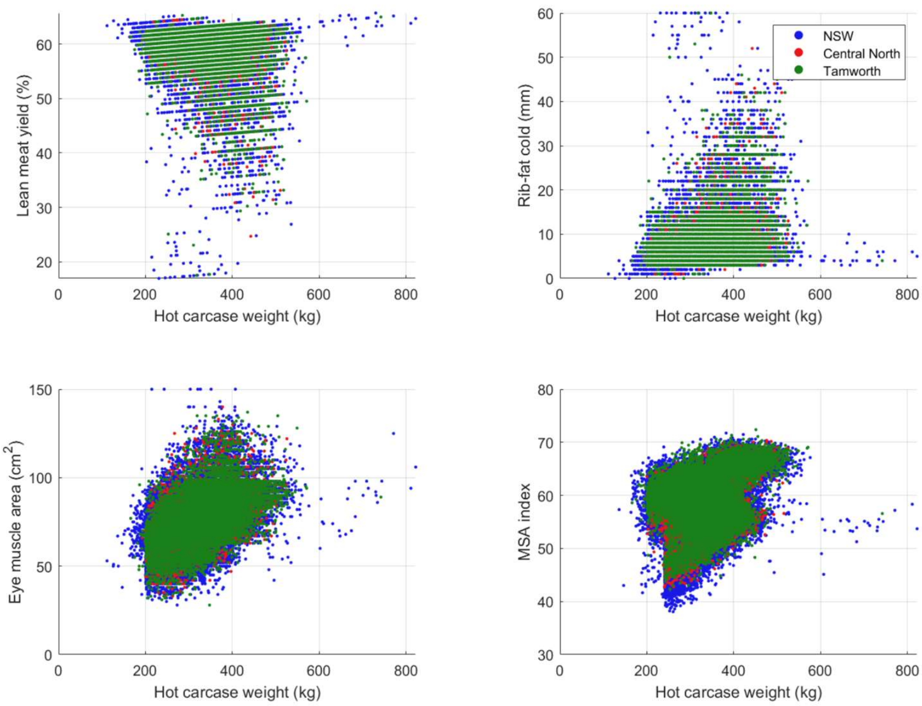

Figure 32. As for the previous figure, but with carcases from the Tamworth shire identified.

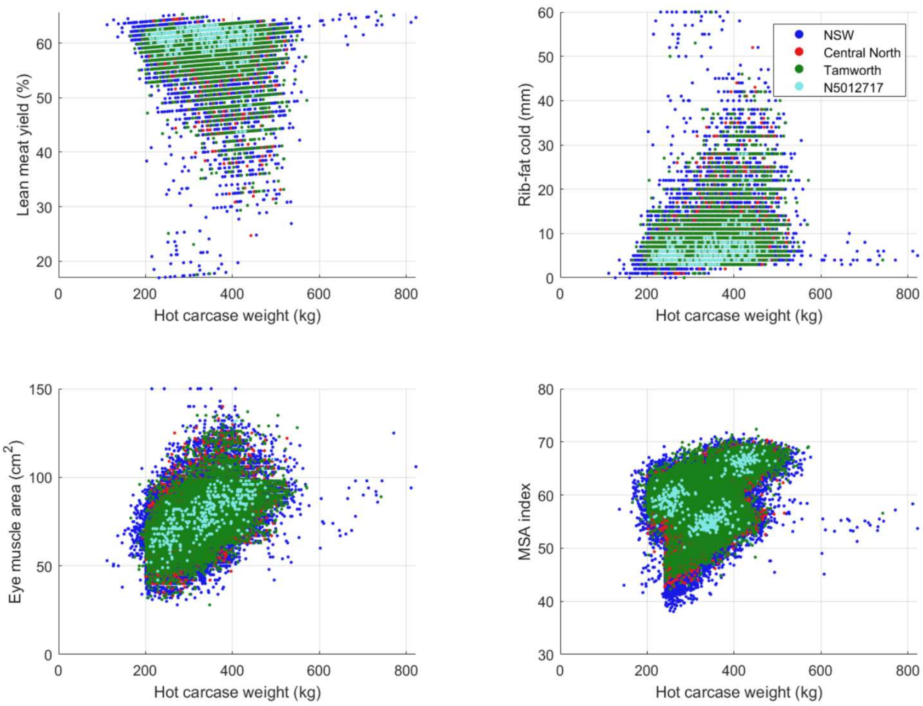

Figure 33. As for the previous figure, but with carcases from an individual producer (PIC) identified.

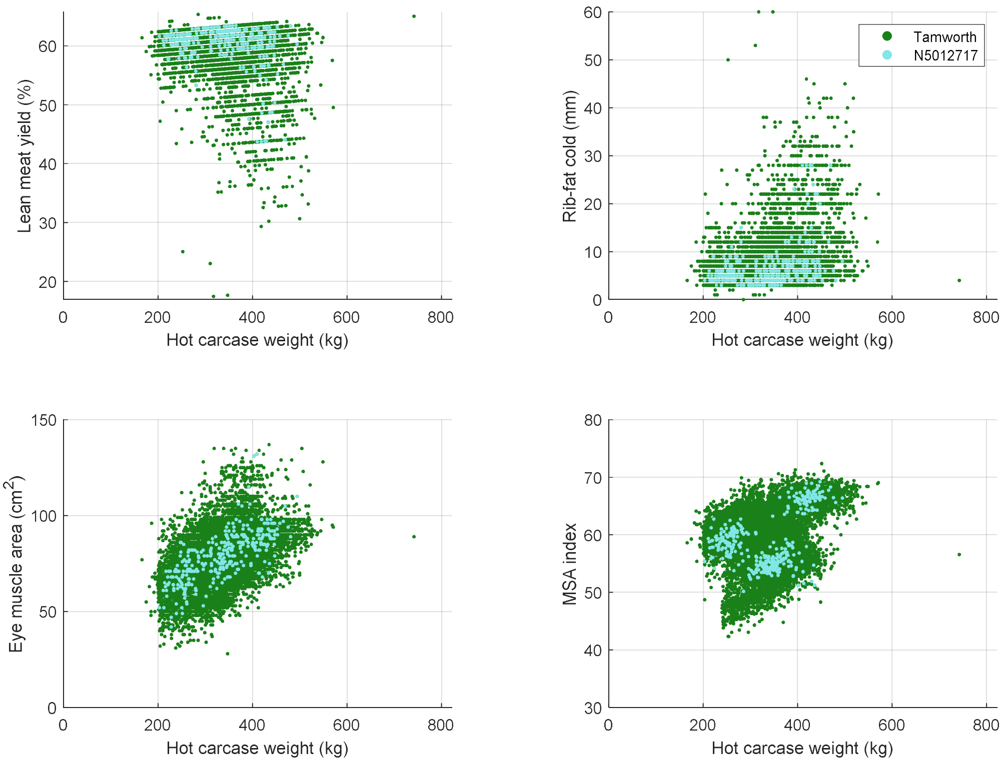

Figure 34. These plots present only the Shire and PIC level data from the previous figure.

Note that in Figure 34, the axes on each plot retain the same scale (range) as was used on the previous figures. This should make visual comparison and interpretation easier for the user.

The following figures present a similar geographical hierarchy, but as single scatter-histogram plots. As previously noted, this format has the advantage over scatter plots of showing the frequency distribution of points along both the X and Y axes. For the example presented in these three figures, the carcases for the chosen PIC are lighter in weight and have a lower MSA index than those of the shire, region and State, but there is a lot of overlap.

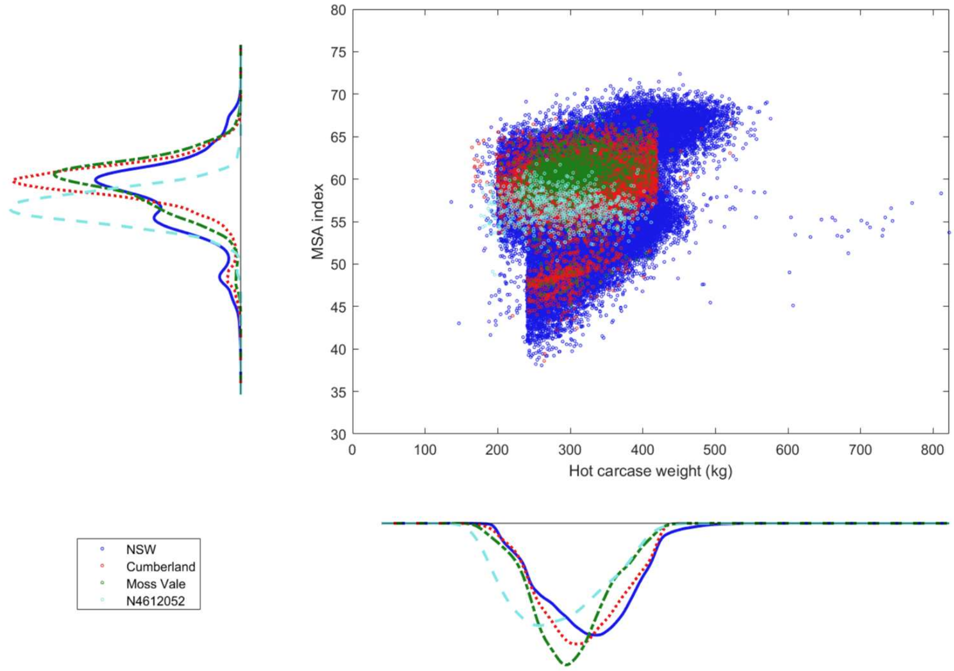

Figure 35. MSA index versus carcase weight colour coded for State (NSW), region (Cumberland), shire (Moss Vale) and a breeder PIC.

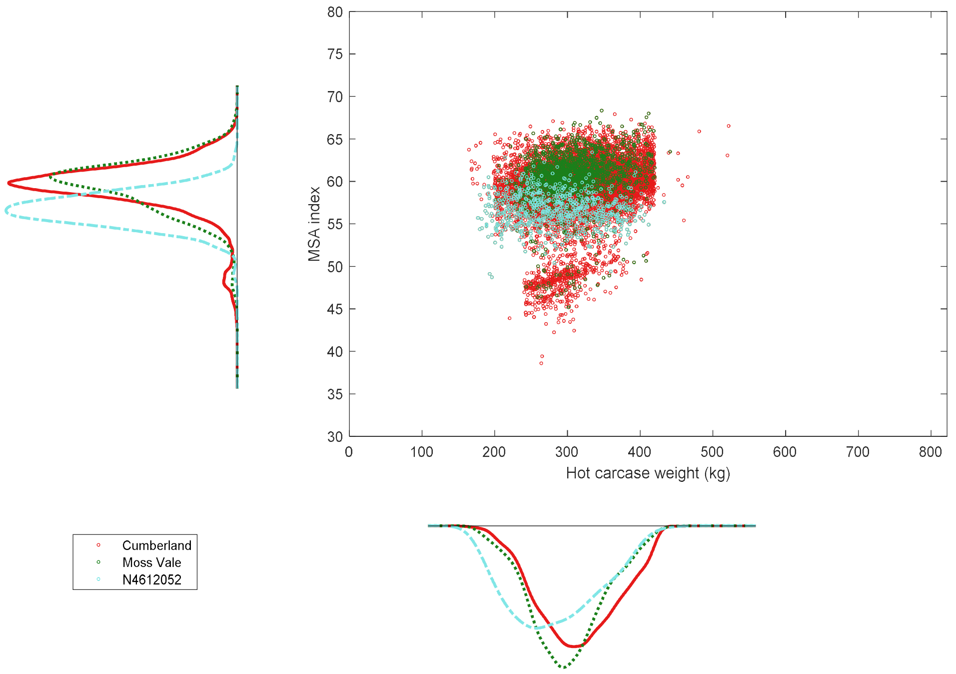

Figure 36. This figure shows the same data as is presented in the previous figure minus the State level data.

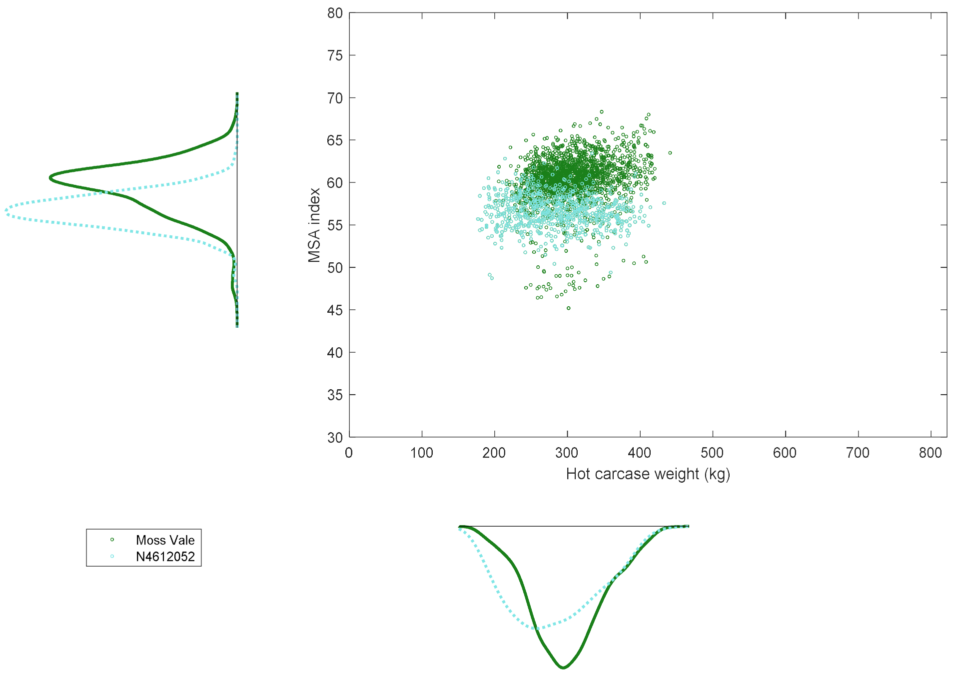

Figure 37. This figure shows the same data as is presented in the previous figure minus the region level data.

The next example shows data for carcases sourced from Queensland and highlights a PIC that produces carcases that mostly have a higher MSA index than those from other producers represented in LDL from the same shire, region and state, but where the carcase weight distribution is like that for the shire, region and State.

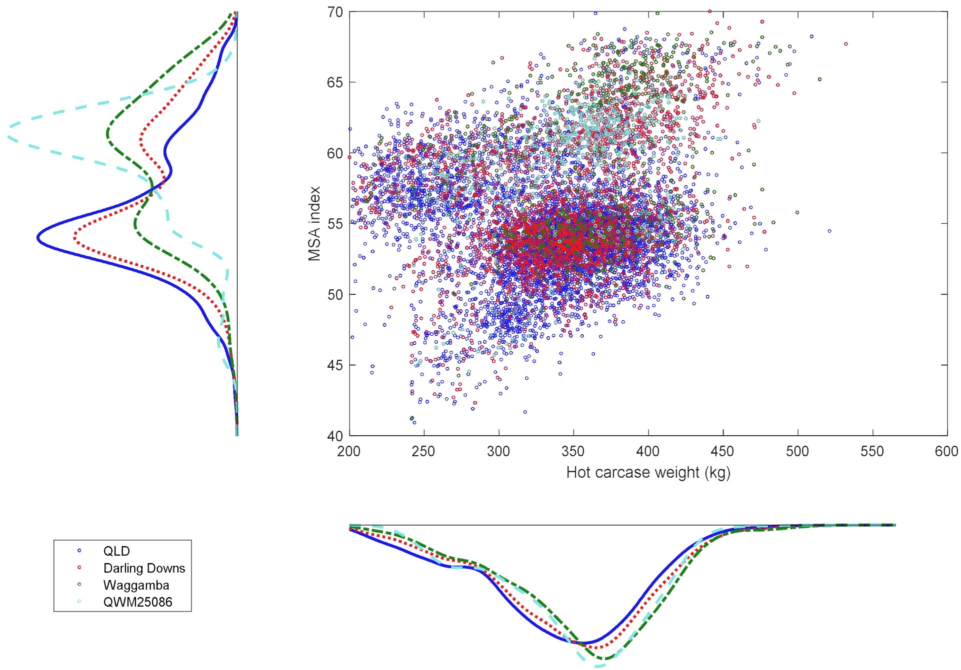

Figure 38. MSA index versus carcase weight colour coded for State (Qld.), region (Darling Downs), shire (Waggamba) and a breeder PIC.

5.2.7 Comparison of aggregates (insights)

Beyond geographical benchmarking, a producer might find value in looking at how the measured variables (e.g. carcase weight, Rib-fat depth, MSA index) vary with different values of the factors recorded for the carcase. As has been listed earlier in this report, these factors include sex, dentition, feed type (grass versus grain), HGP use, breed etc.

The following examples include figures where factors are drawn separately and where multiple levels of the same factor are drawn on the one figure but coded by colour. This latter comparison by colour works for understanding the impact of that factor but cannot be easily used to benchmark a producers' carcases with those from other producers. For producer benchmarking, the figure needs to be reduced to a single level of a factor e.g. sex = male, dentition = 0, and then the producers' carcases and the other carcases can be compared.

5.2.7.1 Sex

The following figures provide examples of possible comparisons. In the first figure, Figure 39, that compares females and males, there is a lot of overlap resulting in a large proportion of the females being hidden. While the side frequency distribution helps to understand the spread, a better option would be to allow the user to show just one sex at a time, or to be able to reverse the order of display. Figure 40 presents just the female data showing that even though there is a dip in the MSA index frequency distribution at an index of approximately 52, there are carcases covering the same carcase weight range as at both lower and higher index values.

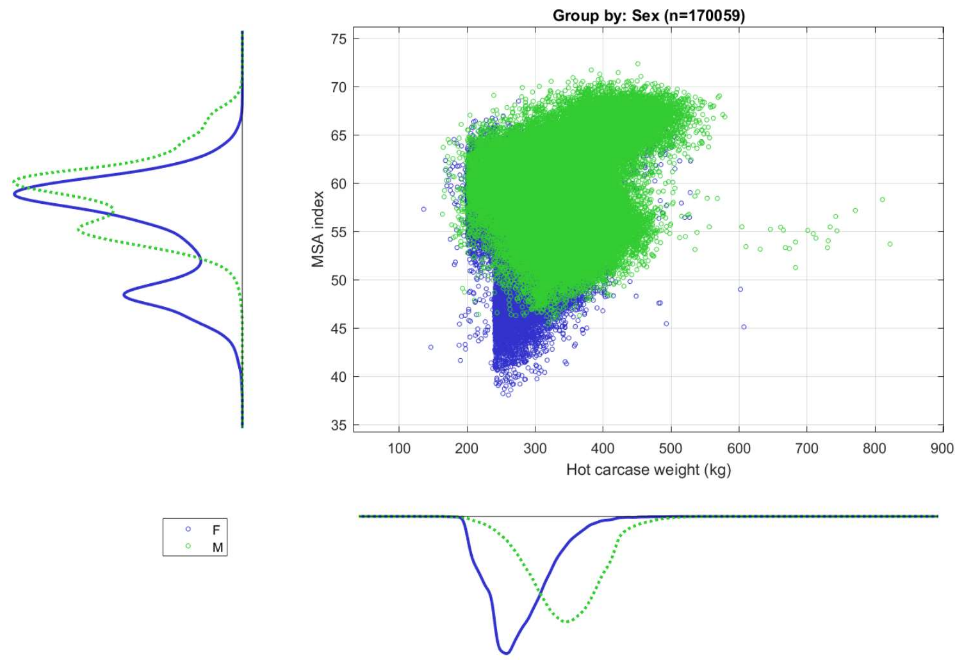

Figure 39. Comparison of the distribution of carcases on an MSA index by carcase weight surface comparing males and females.

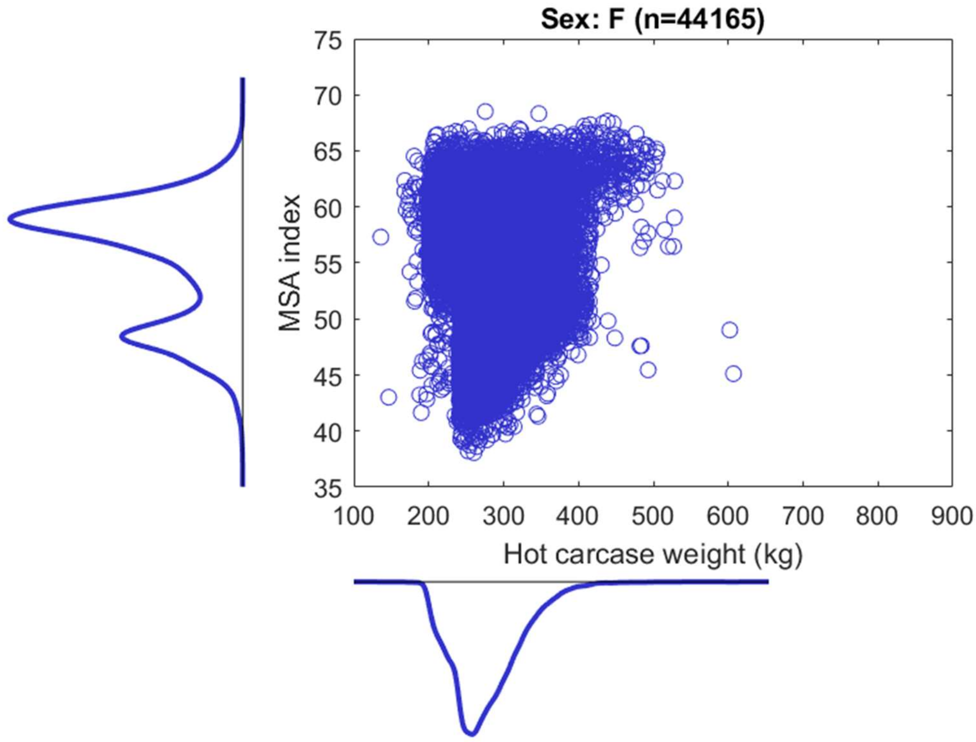

Figure 40. Distribution of carcases on an MSA index by carcase weight surface, females only.

5.2.7.2 Dentition

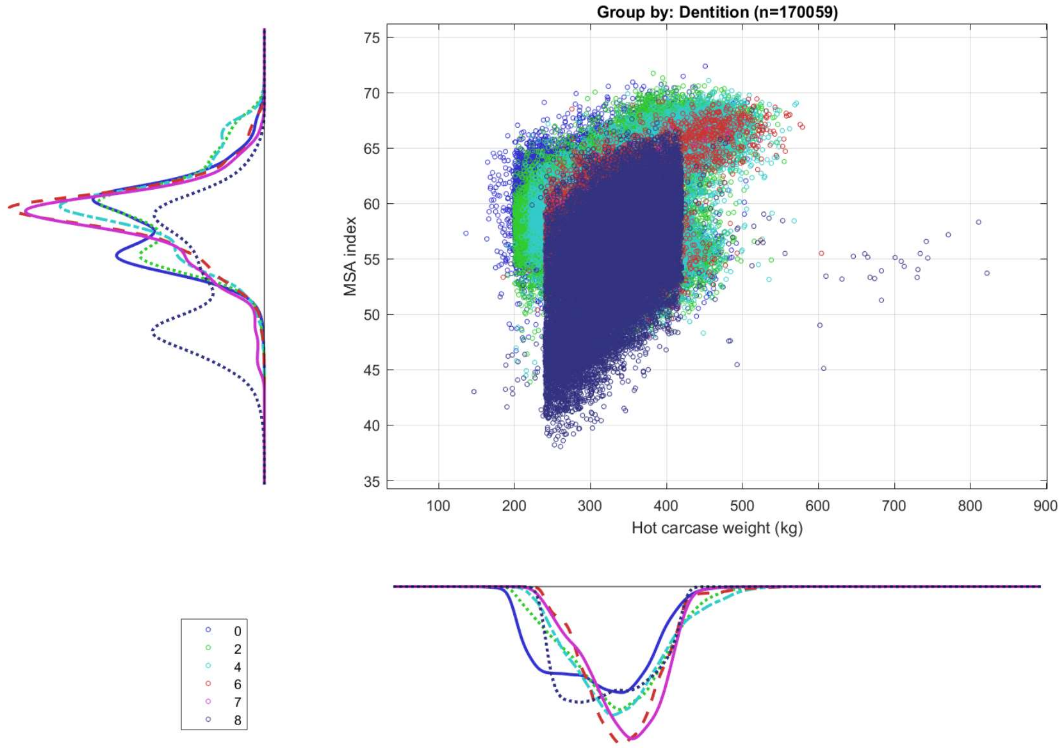

Figure 41. Distribution of carcases on an MSA index by carcase weight plane colour coded by dentition.

Figure 41 shows the distribution of carcases by dentition, except that the large number of carcases with a dentition of '8' masked a large area of the surface. As a result, this is one factor that needs to be drawn separately for each dentition level, or the user needs to be able to choose which levels to display. The following figures show the carcase distribution for four individual dentition levels (0, 2, 4, 8).

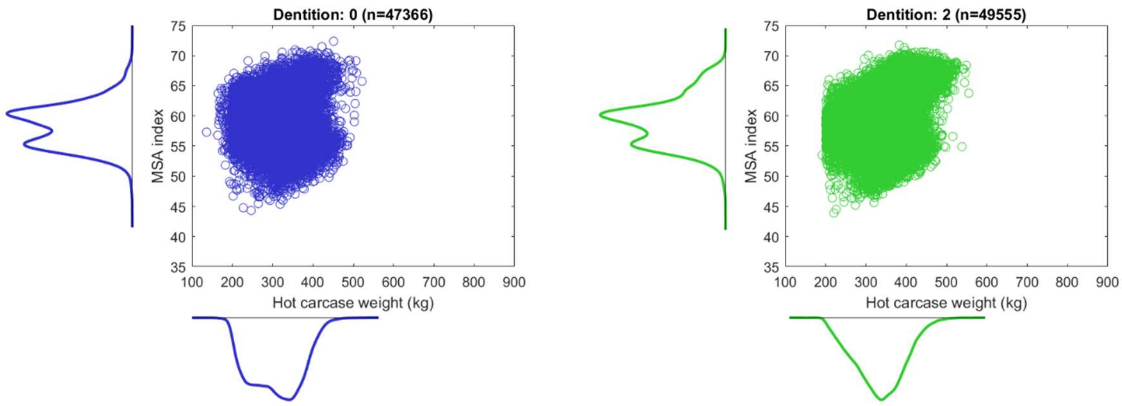

Figure 42. Distribution of carcases with a dentition level of '0' (left) and '2' (right) on an MSA index by carcase weight plane.

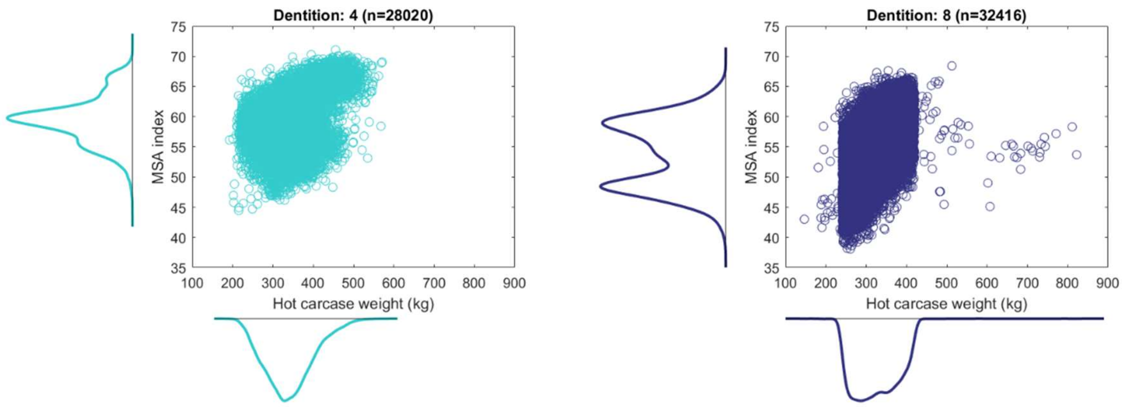

Figure 43. Distribution of carcases with a dentition level of '4' (left) and '8' (right) on an MSA index by carcase weight plane.

5.2.7.3 Hormonal growth promotant (HGP) use

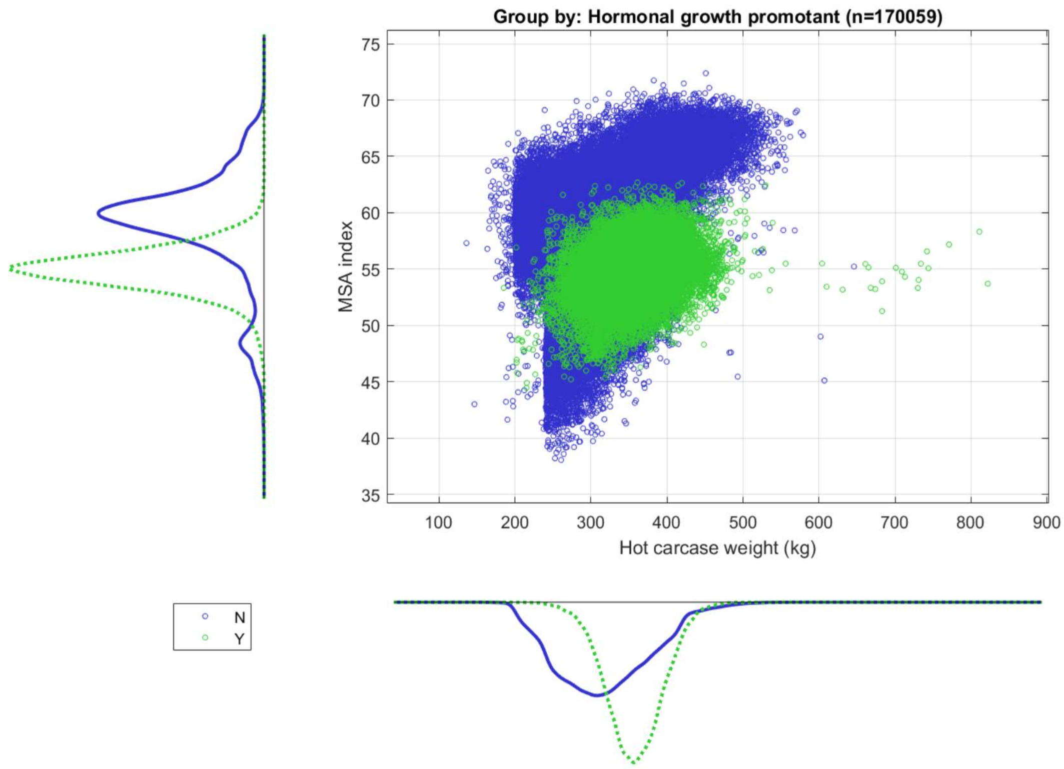

Figure 44. HGP used in green versus not used in blue on the MSA index by carcase weight surface.

This figure (Figure 44) comparing HGP use against carcases where HGPs were not used indicates that on average the HGP cattle were heavier, but they did not have as a high an MSA index as most of the equivalent weight carcases that were not produced with HGPs – HGP use has a negative impact on the MSA index (5 points) and this is apparent in the data.

5.2.7.4 Grain versus Grass finished

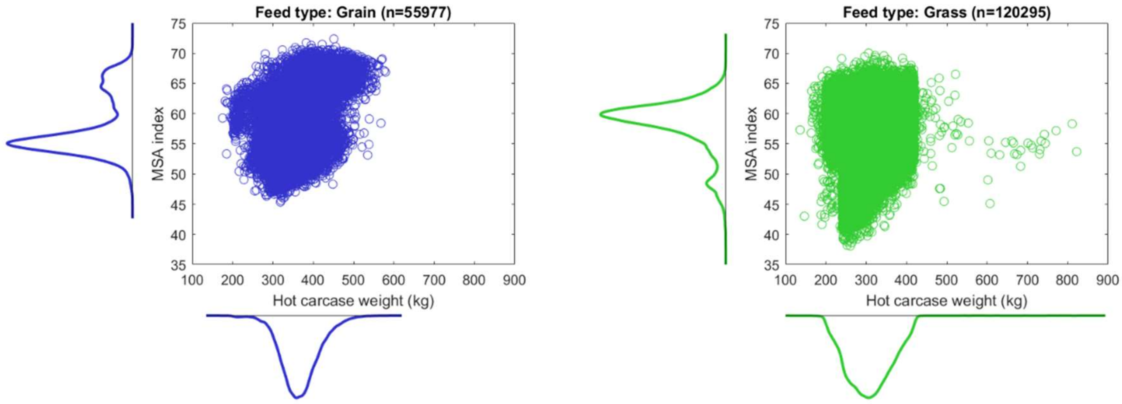

Figure 45. Distribution of carcases on an MSA index by carcase weight plane either grain finished (left chart) or grass finished (right chart).

Grain versus grass finished is another comparison best separated due to the heavy overlap. These charts seem to indicate higher MSA index values and a wider range of carcase weights for the grain finished carcases. The low MSA index carcases (less than 45) were all grass finished and likely to be older animals.

5.2.7.5 AUS marbling

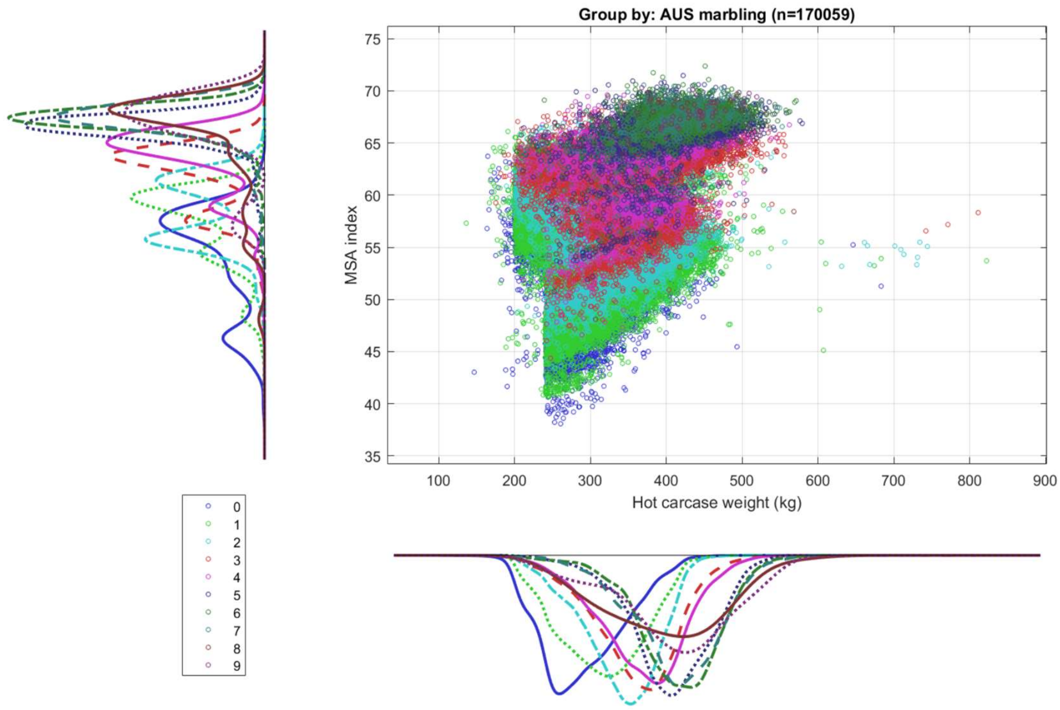

Figure 46. The complex layers of carcases on an MSA index by carcase weight plane coloured by AUS marbling score.

In general, MSA index increases with increasing carcase weight and increasing marbling score. This can be seen in the following composite figure (Figure 47). This insight is important for producers as it shows the correlation between the MSA index and carcase weight within different marble scores and potentially highlights some of the discrepancies that occur in grids that have both an MSA and an AUSMEAT marble score target.

This is an example where including all AUS marbling levels on one plot (Figure 46) makes it too complex. In drawing each level separately, it is crucial that the same axis scales are used on each plot.

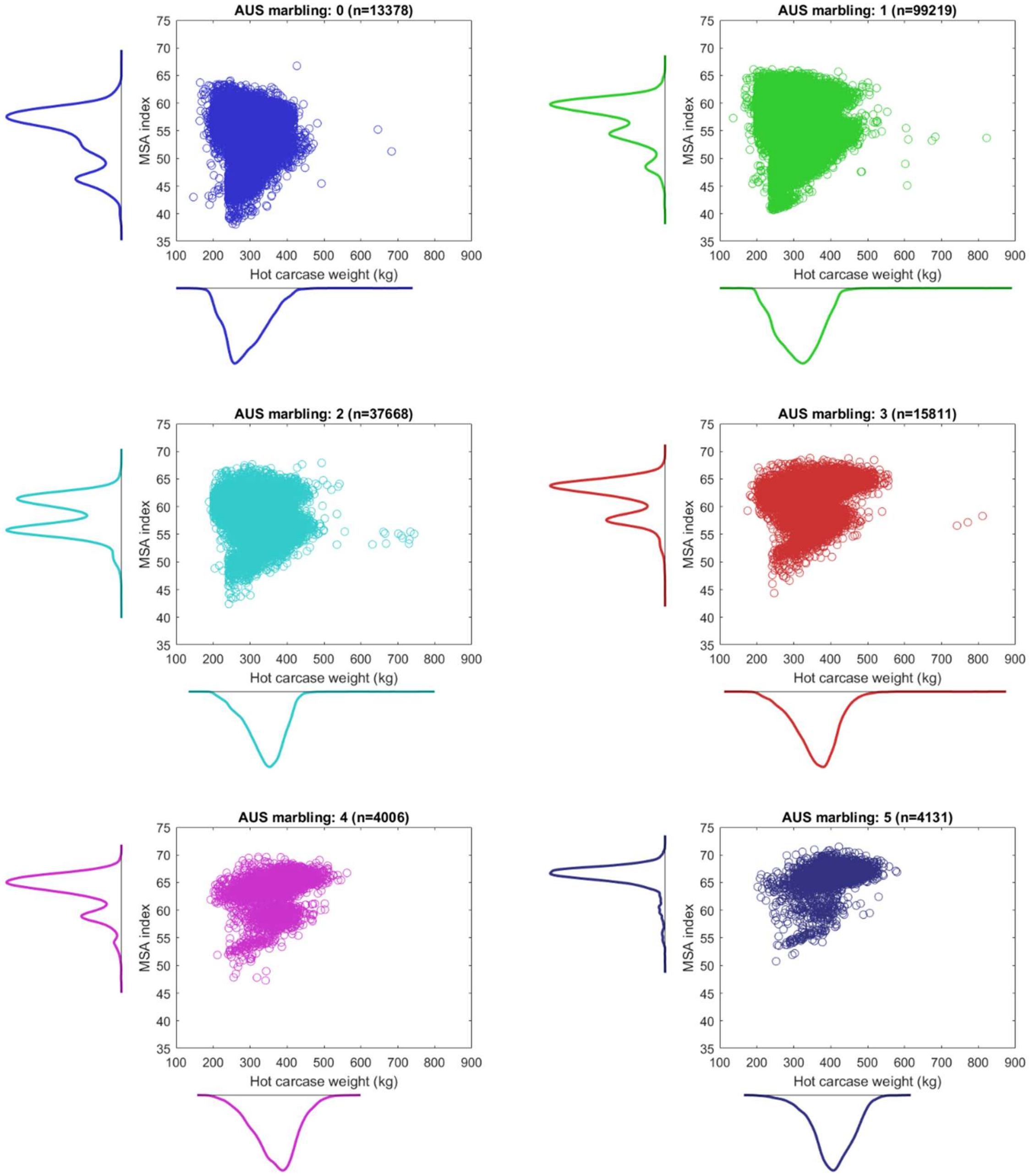

Figure 47. Comparison of the distribution of carcases for different AUS marbling scores.

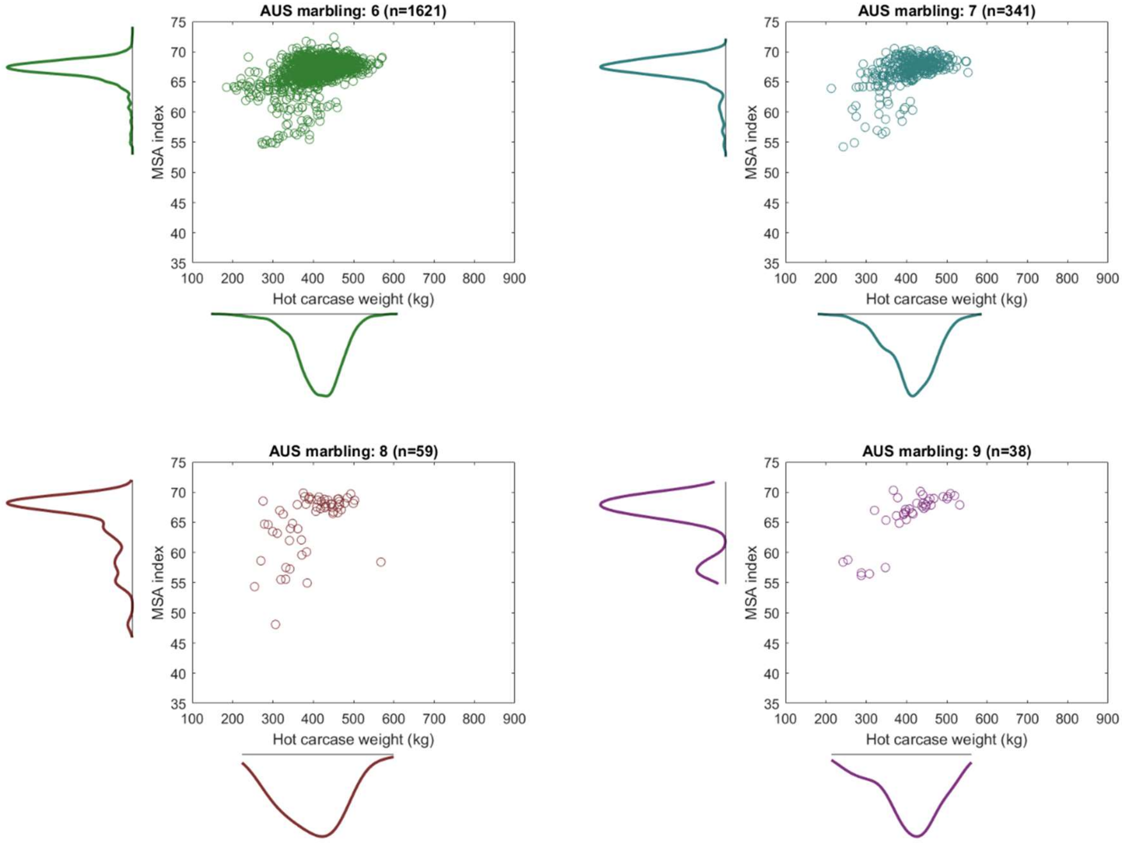

Figure 47. Comparison of the distribution of carcases for different AUS marbling scores. (continued)

Note that where there are few data points, for example the plots for AUS marbling '8' and '9', the side histograms are less useful but also less important as the density of points can be seen in the scatter plot.

5.3 Recommendations on aggregation formats

This report has examined several standard graphical techniques for displaying combinations of variables and factors (aggregates) within the LDL database. The advantages and disadvantages of these graphical interfaces are given for each type. After extensive internal review, the project team (based on significant industry experience) recommends that for a single trait or factor a combination of a table and line graph would be appropriate given that producers and processors are more likely to require this format to be plotted against time. For situations where two or more traits or factors need to be visualised a combination of scatterplots and histograms should be used as this provides necessary information on the variance, range and potential outliers.

These combinations provide the user (a producer) with the best visualisation of the relationships between the components, the density of those components and the distribution of each of the components. This combination also allows for multiple levels of factors (aggregates) to be overlaid which provides the producer with the aggregation levels that were identified in the survey to be important.

6 Discussion

The key objective of LDL is to construct a resource that has the capacity to translate data collected on carcase attributes into information that can be used in decisions that improve compliance to market specifications and reduces the costs of non-compliance. As shown in the preceding sections the LDL database contains an extensive set of carcase traits that can be aggregated into various insights based on time (kill date and season), production type, market type and location (local, regional, state and national). All these insights provide single points of reference for producers and processors that are best visualised with a combination of scatterplots and histograms. This combination provides the best graphic representation of distributions, correlations, ranges, medians and outliers.

However, as many of these traits have strong underlying genetic and phenotypic correlations or relationships, presenting the insights on single traits or combinations of two traits can often miss vital interactions between traits (factors). For this reason, the project team has recommended that LDL consider the development of insight dashboards which are described in the next section.

6.1 LDL Insights dashboard concepts

In the following section, exemplars of LDL insight dashboards have been created. These dashboards have been created based on an investigation of the structure of the LDL dataset, the fields that are available for aggregation and feedback from producers in the LDL survey. In all cases, the LDL cattle dataset was used as it provided a more comprehensive and complex set of factors and insights that could be modelled.

In designing the structure of the dashboard, an important consideration was providing the ability for the end-user to select the required insights from a series of “buttons” or fields. The first example (Figure 48) shows a simple combination of the tabular format for cattle supply and carcase weight (count, mean, variation and distribution) for the current year, the previous year and averages for the previous three years. Below the table is a plot of the average carcase weights of those cattle by time.

The measured traits shown as tabs or buttons across the top of the dashboard can be selected to change the trait displayed. Figure 49 shows the corresponding data for Rib-fat depth.

The end-user can then select a trait of interest to further filter the analysis displayed in the body of the dashboard. These “filters” are shown as buttons/tabs down the left side of the dashboard. Figure 50 provides an example where the data is filtered to only include carcases labelled “Male”.

HSCW | Rib-fat | LMY% | MSA index

All
Sex
Dentition
HGP use
AUS marbling

User filter

Region: Mid-Coast	Statistics	Current year (2018-19)	Previous year (2017-18)	Previous 3 years
PIC: Nx26nnn4	Number of carcases	1,692	597	766
	Mean (std.dev.)	322 (47.8)	310 (68.2)	294 (59.8)
	5^{th},95^{th} percentile	228 – 416	176 – 444	177 – 411
Shire: Gloucester	Number of carcases	27,703	23,066	25,561
	Mean (std.dev.)	270 (82.4)	280 (86.8)	261 (84.6)
	5^{th},95^{th} percentile	108 – 431	109 – 450	95 - 426
Region: Mid-Coast	Number of carcases	51,650	45,836	53,141
	Mean (std.dev.)	266 (85.2)	281 (89.0)	257 (90.5)
	5^{th},95^{th} percentile	99 – 433	107 – 456	80 – 435
State: NSW	Number of carcases	187,345	160,314	206,345
	Mean (std.dev.)	263 (100)	273 (101)	243 (100)
	5^{th},95^{th} percentile	67 – 459	75 – 472	47 - 439

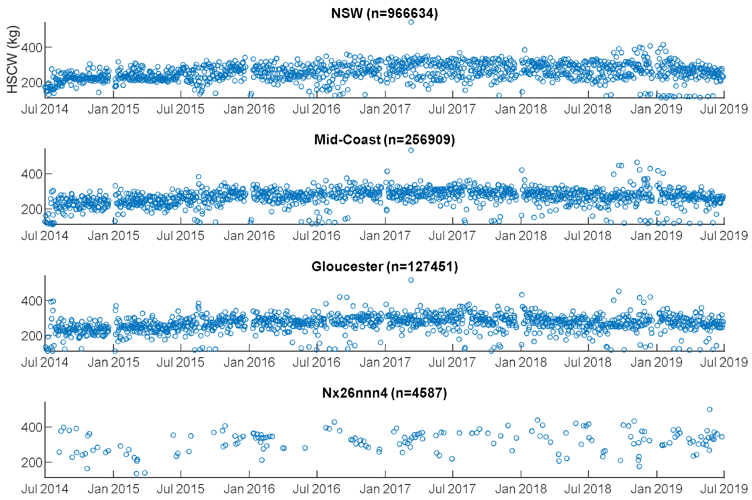

Figure 48. Dashboard showing aggregate statistics for cattle supply and carcase weight for current year, last year and average for previous three years, and time series charts of average carcase weight per kill day by producer through to state.

Region: Mid-Coast	Statistics	Current year (2018-19	Previous year (2017-18)	Previous 3 years
PIC: Nx26nnn4	Number of carcases	261	137	138
	Mean (std.dev.)	4.8 (1.8)	4.6 (2.3)	4.7 (1.9)
	5th,95th percentile	1.2 – 8.3	0* - 9.2	1.0 – 8.4
Shire: Gloucester	Number of carcases	8,108	7,955	5,033
	Mean (std.dev.)	4.8 (2.8)	4.6 (2.4)	4.9 (2.2)
	5th,95th percentile	0* - 10.4	0* - 9.4	0* - 9.2
Region: Mid-Coast	Number of carcases	14,867	15,436	10,937
	Mean (std.dev.)	4.9 (2.9)	4.7 (2.4)	5.1 (2.3)
	5th,95th percentile	0* - 10.6	0.0 – 9.3	0.6 – 9.6
State: NSW	Number of carcases	30,719	34,628	29,515
	Mean (std.dev.)	5.8 (4.1)	5.1 (3.2)	5.5 (2.7)
	5th,95th percentile	0* - 13.9	0* - 11.3	0.1 – 10.8

Figure 49. Dashboard for trait by production characteristic x time series. Rib-Fat selected.

Region: Mid-Coast	Statistics	Current year (2018-19)	Previous year (2017-18)	Previous 3 years
PIC: Nx26nnn4	Number of carcases	999	320	358
	Mean (std.dev.)	349 (28.1)	363 (25.1)	338 (40.3)
	5^{th},95^{th} percentile	294 – 404	314 – 413	259 – 417
Shire: Gloucester	Number of carcases	11,958	11,051	11,276
	Mean (std.dev.)	318 (76.4)	331 (71.3)	311 (75.3)
	5^{th},95^{th} percentile	168 – 468	192 – 471	163 - 458
Region: Mid-Coast	Number of carcases	21,860	22,566	23,862
	Mean (std.dev.)	321 (75.4)	338 (66.0)	320 (67.7)
	5^{th},95^{th} percentile	173 – 469	209 – 467	187 – 452
State: NSW	Number of carcases	77,939	75,521	82,246
	Mean (std.dev.)	336 (73.2)	343 (66.2)	322 (67.4)
	5^{th},95^{th} percentile	193 – 480	213 – 473	190 – 454

Figure 50. Dashboard of trait x production type x time series. Carcase weight selected. Sex selected to Male

The insights dashboard can then change from the tabular/histogram combination that best describes a single trait, to a second dashboard that combines two traits, whilst still retaining the functionality of selection of trait type and also production characteristic as shown in Figure 51.

In this example, MSA index is plotted against carcase weight (both selected from the tabs at the top of the dashboard) and then colour coded for sex (selected from the left-side tabs). The combination of overlaid scatter plots and side-bar histograms shows the spread or extent of the measures as well as the difference in the distributions between the sexes.

This type of dashboard could also have user defined thresholds for traits added which are consistent with market specifications.

The following example, Figure 52, presents scatter plots of the same variables (MSA index by carcase weight), but separated into individual charts for each state of origin of the carcase. The LDL data allows the option to choose “state of origin” either based on the producer that delivered the carcase to the processor or the breeder of the animal.

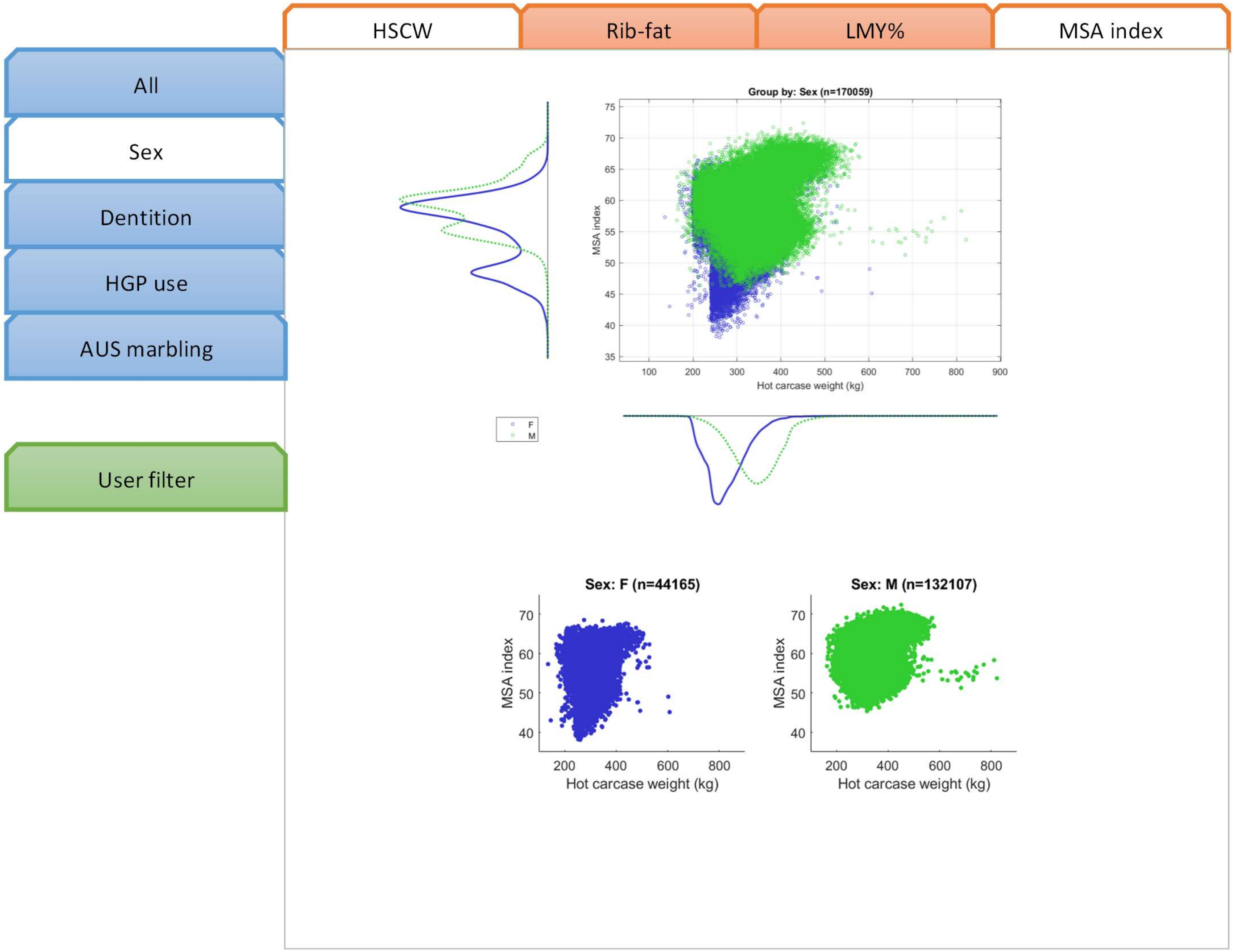

Figure 51. Dashboard of a scatterplot / histogram combination that allows the end-user to visualise two traits (carcase weight and MSA index) that have a correlation and the distribution of those traits by a defined production characteristic (Sex).

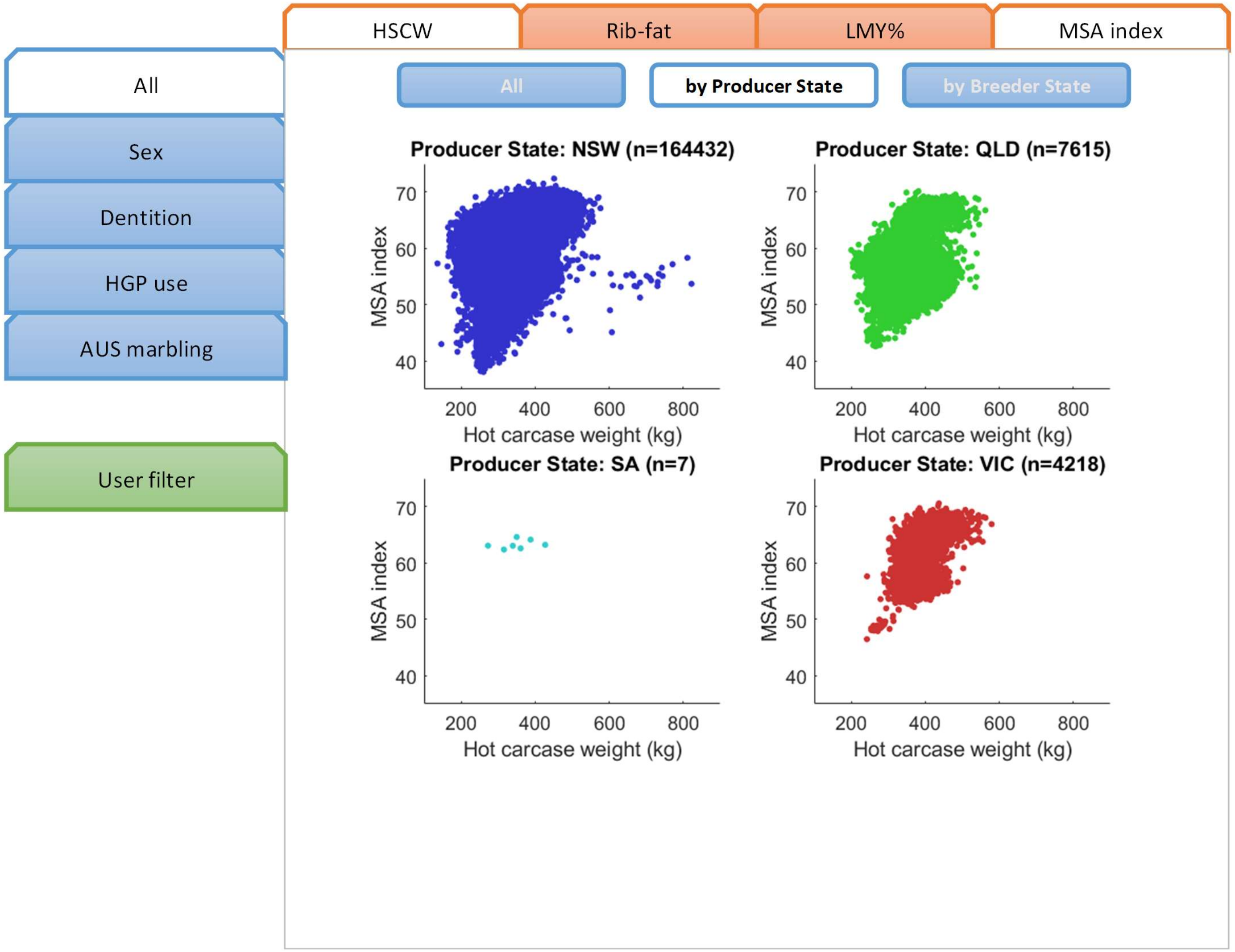

Figure 52. Dashboard containing a grid of scatter plots that allows the end-user to visualise two traits (carcase weight and MSA index) that have a correlation and the distribution of those traits by a defined production characteristic (Producer state).

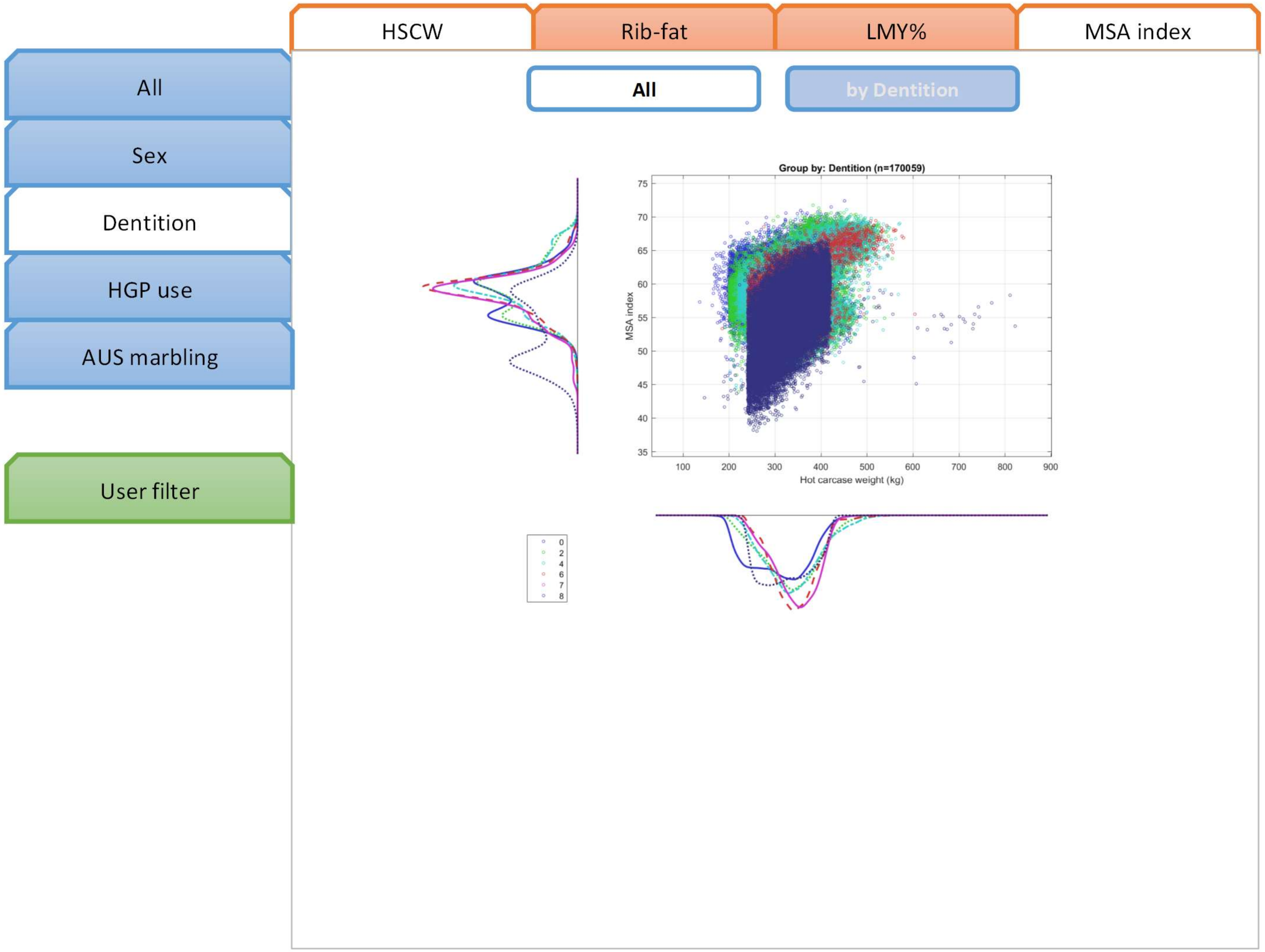

Figure 53. Dashboard showing a scatter plot of carcase weight by MSA index coloured by Dentition. Side-bar plots show the frequency distribution for each dentition level.

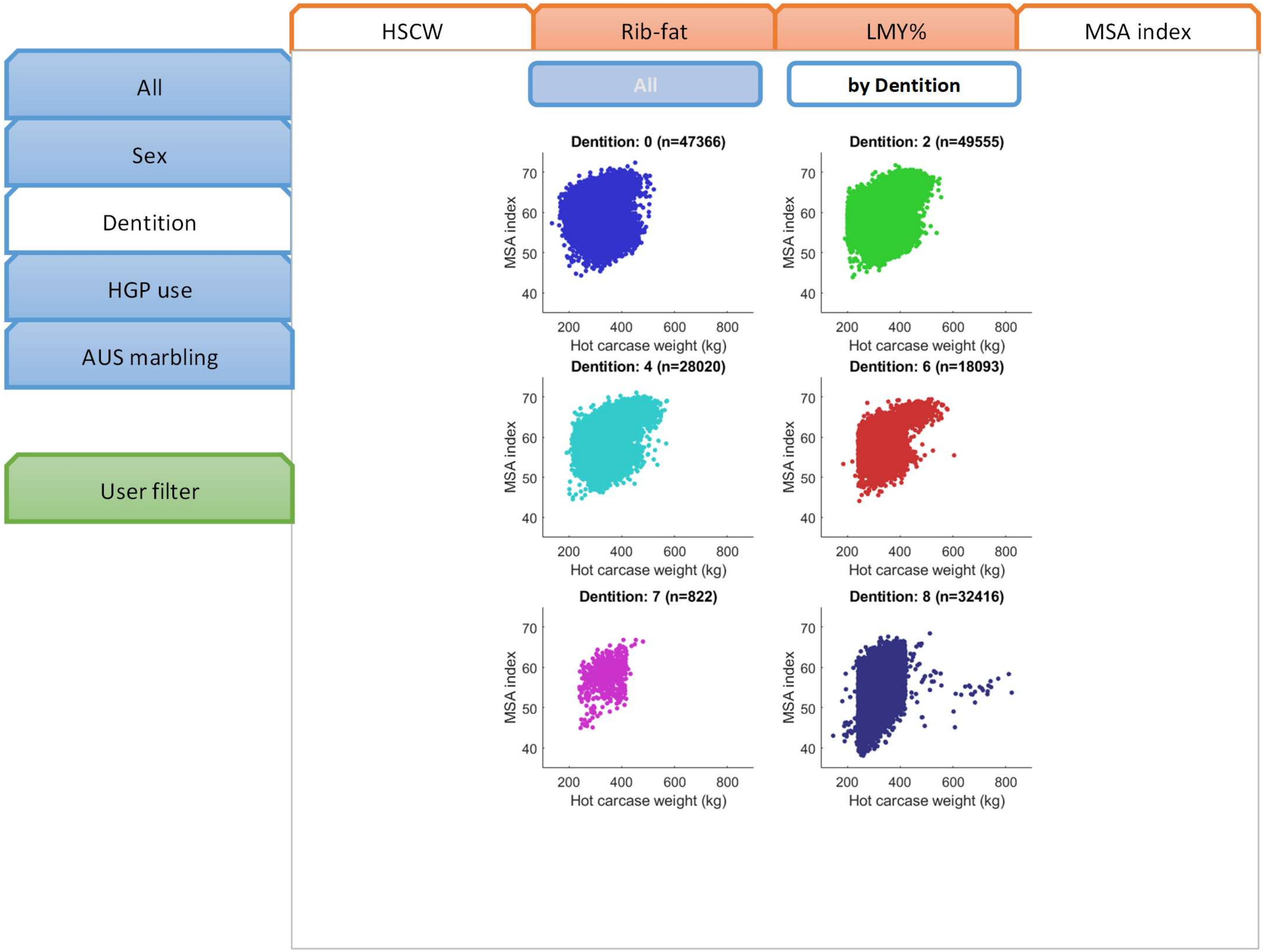

Figure 54. Scatter plots showing the relationship between carcase weight and MSA index separated into individual plots by dentition level.

Dentition could be an important attribute in some markets. Figure 53 shows a scatter/histogram plot of MSA index versus carcase weight coloured by dentition. Where there is a large number of carcases as in this example, the later drawn dentition levels often obscure the first drawn levels. To overcome this, the user could be given the option to draw separate scatter plots for each dentition level (Figure 54).

Another option for exploring the relationship between a measured trait (e.g. carcase weight) and dentition is to provide a series of pie charts. In Figure 55, separate pies have been drawn for the user's PIC and the corresponding shire, region and state. The size of each slice of a pie indicates the proportion of carcases from that area that had a specific dentition level, and each slice is labelled with the average of that measured trait, in the example, carcase weight.

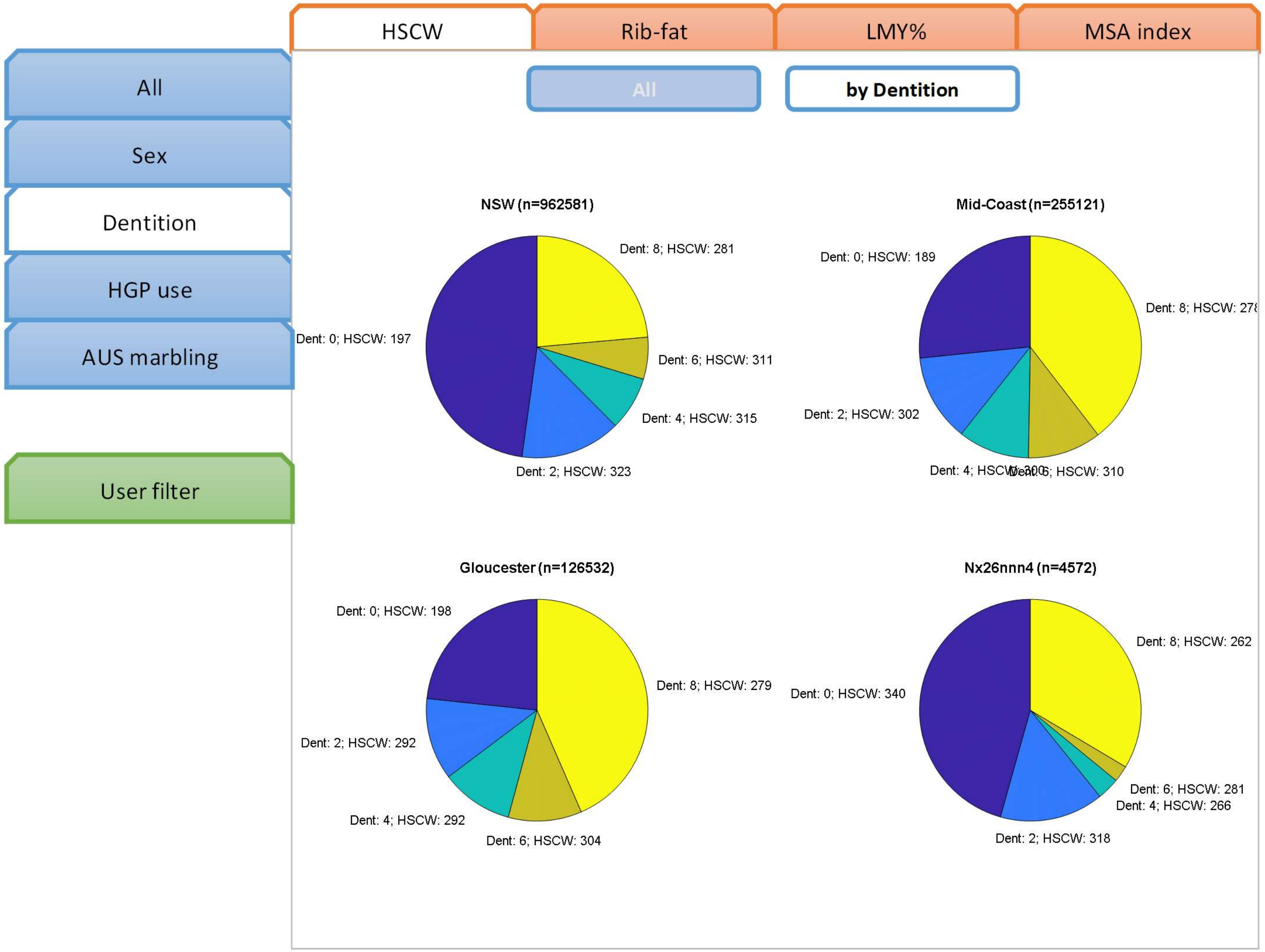

Figure 55. Pie charts show the distribution of carcases by dentition level for the user's PIC, and the corresponding shire, region and state. The average carcase weight is reported with each area x dentition combination (each 'slice of a pie').

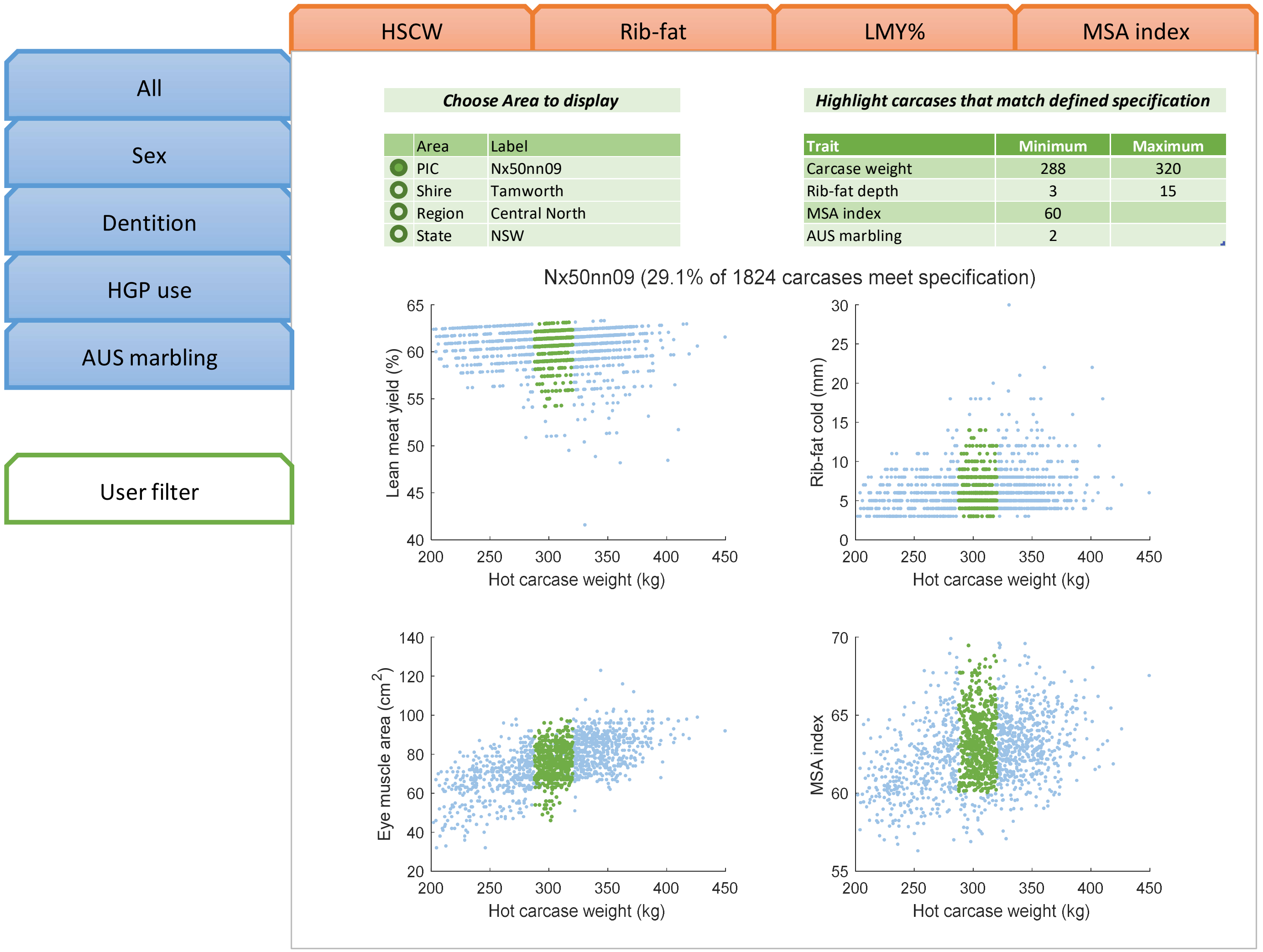

Figure 56. This example dashboard shows scatter charts of carcase weight versus each of lean meat yield %, Rib-fat depth, Eye muscle area and MSA index. Carcases that match a user defined specification are highlighted.

A key benefit of using an interactive resource such as LDL, is the opportunity to allow the user to customise the output to their specific interests. This is not limited to seeing how their carcases compare to others from the same region or between periods but can show which ones meet a user chosen specification.

Figure 56 shows an example where all carcases for a user's PIC are shown on scatter plots of carcase weight against lean meat yield %, Rib-fat depth, Eye muscle area and MSA index. Above the charts are two panels where (with the left-side panel) the user can select which carcases to display – those from the user's PIC or from the corresponding shire, region or state. The right-side panel is used to define a specification, ranges for carcase weight, Rib-fat depth, MSA index and AUS marbling in this example. Those carcases that match the specification are identified in a different colour, green in this example. The proportion of compliant carcases as well as the total number is shown in the title directly above the charts.

The following dashboard, Figure 57, shows the same layout and specification, but with the region selected, Central North in this example. It is interesting to note that only 1.4% of carcases from the Central North region matched the example specification, compared to 29.1% of carcases from the example PIC.

A final example, Figure 58, uses box plots to show the range of MSA index values for different AUS marbling scores. In the example provided separate plots have been shown for the user's PIC and the corresponding shire, region and state. The upper and lower extents of the blue boxes represent the 75^{th} and 25^{th} percentiles respectively with the median shown as the red bar within the box. The black whiskers extend to the range of data not considered outliers. Outliers are indicated by red crosses.

These box plots show that there is a generally an increasing trend in MSA index with increasing marbling score up to score 6, though in the example shown almost none of the user's carcases achieved score 6. While MSA index is positively correlated with marbling score, there is a lot of overlap indicating the importance of other traits.

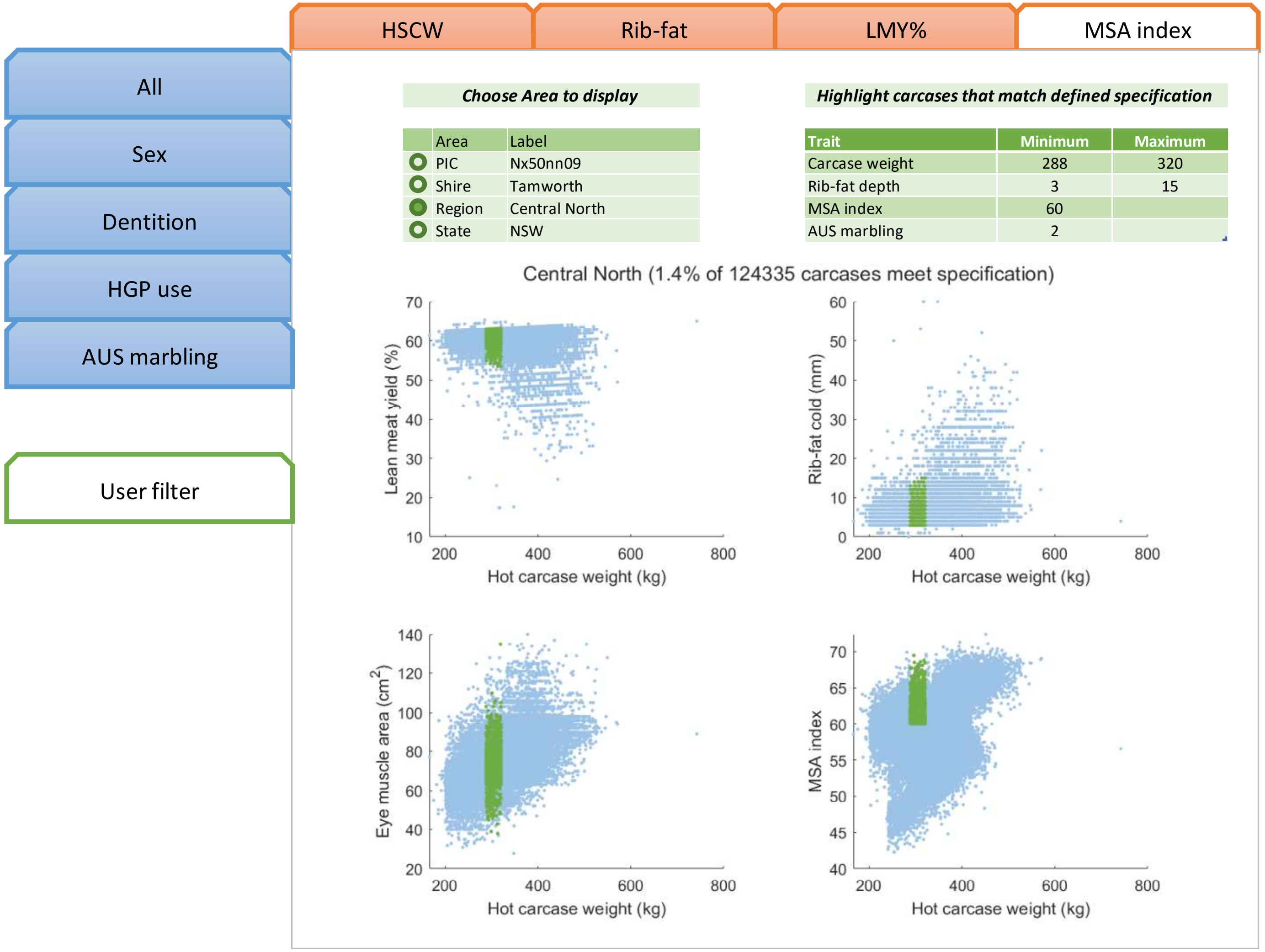

Figure 57. This dashboard has the same layout as the previous figure, shows data for the corresponding region, Central North, NSW.

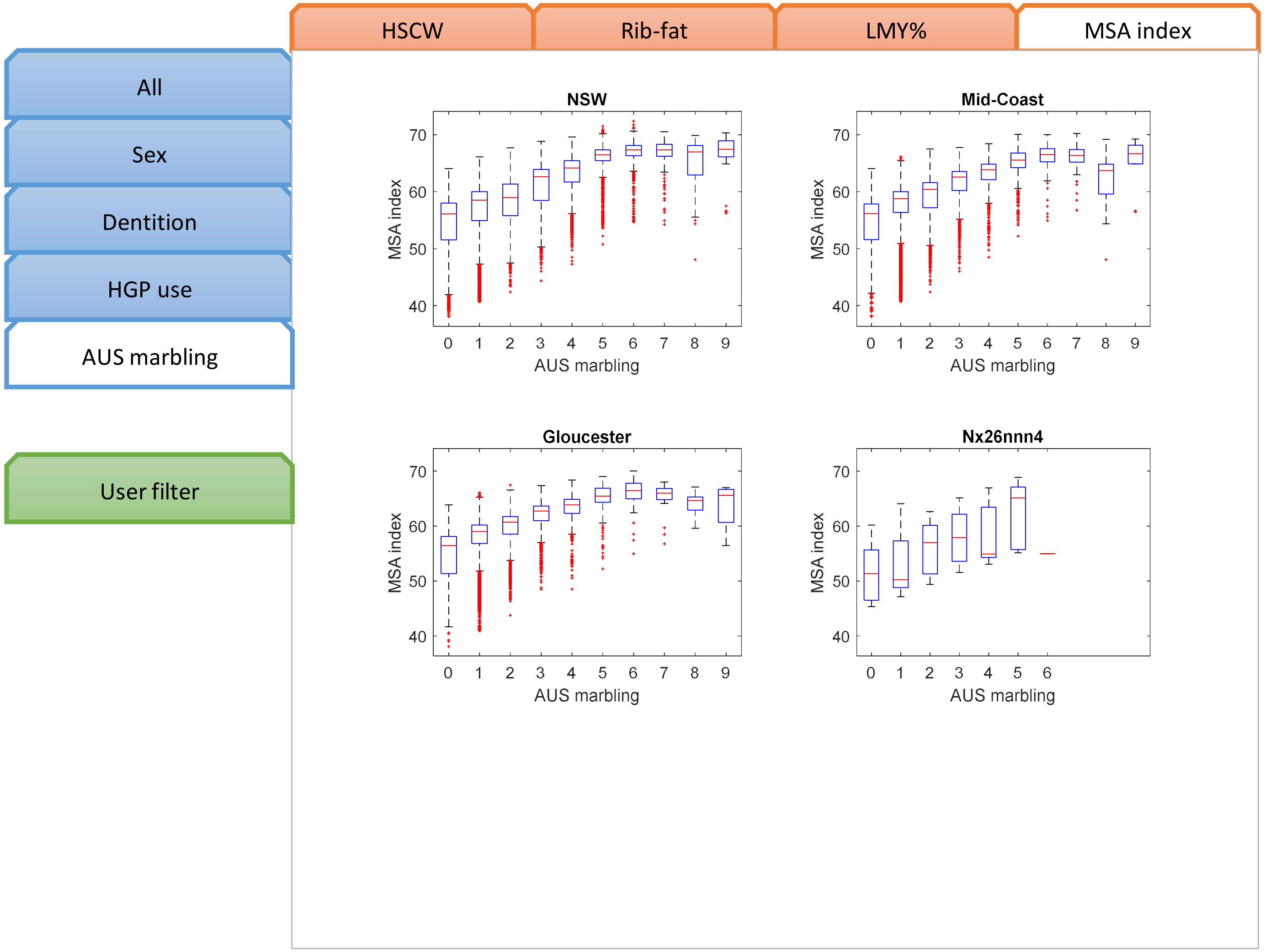

Figure 58. The distribution of MSA index for each AUS marbling score is shown in a boxplot. A separate plot is shown for the user's PIC and for the corresponding shire, region and state.

6.2 Valuing insights in LDL

An objective with this project was to provide some quantification of the value that data insights can provide to producers that are targeting higher compliance through better decision making. One of the key issues with any attempts to put a dollar benefit on the insight, is that the insight provides a retrospective or historical benchmark of the producer's data and may not reflect the next consignments which may have different environmental or production characteristics. However, the LDL insights do enable a visualisation of trends and likely correlations that could be intrinsic properties of future consignments particularly if genetics or feeding environments (days on feed) are maintained.

An example of this shown in Figure 56 where a producer can define a set of market specifications that are aligned to a processor grid or brand. In this case, it clearly shows and is visually apparent that the producer is suppling a high proportion of cattle outside of the grid (70%). A table showing the exact numbers of cattle outside of the grid based on fat, weight and MSA index could be constructed if details on the grid were available, then a cost of that non-compliance could be calculated.

For example, in this grid, animals with a Rib-Fat less than 3 would not be eligible and would attract a discount of 50c. Thus, for an average animal of 310 kg, this translates into $155 in lost value. A heat map of lost value could be constructed. In this case the producer would be able to spend up to $155 dollars to correct the loss by increasing Rib-Fat to greater than 3mm.

For the processor, as shown in Figure 57, being able to graphically display an entire region would be important in determining whether cattle supplied from that region could underpin a sustainable supply to a given set of market specifications. In this case, it is very evident that only a very small proportion (1.4%) of cattle supplied from that region would fit the desired market specifications.

An alternative methodology that should be considered for implementation as a way of demonstrating the lost opportunity for producers and processors is the construction of a grid control chart. In this example a grid that reflects the current purchasing patterns of a major lamb value chain was used. The basis of this grid is shown in Figure 59, which has a discount for carcases above and below a range of 20-24kg and a discount for low and high fat scores.

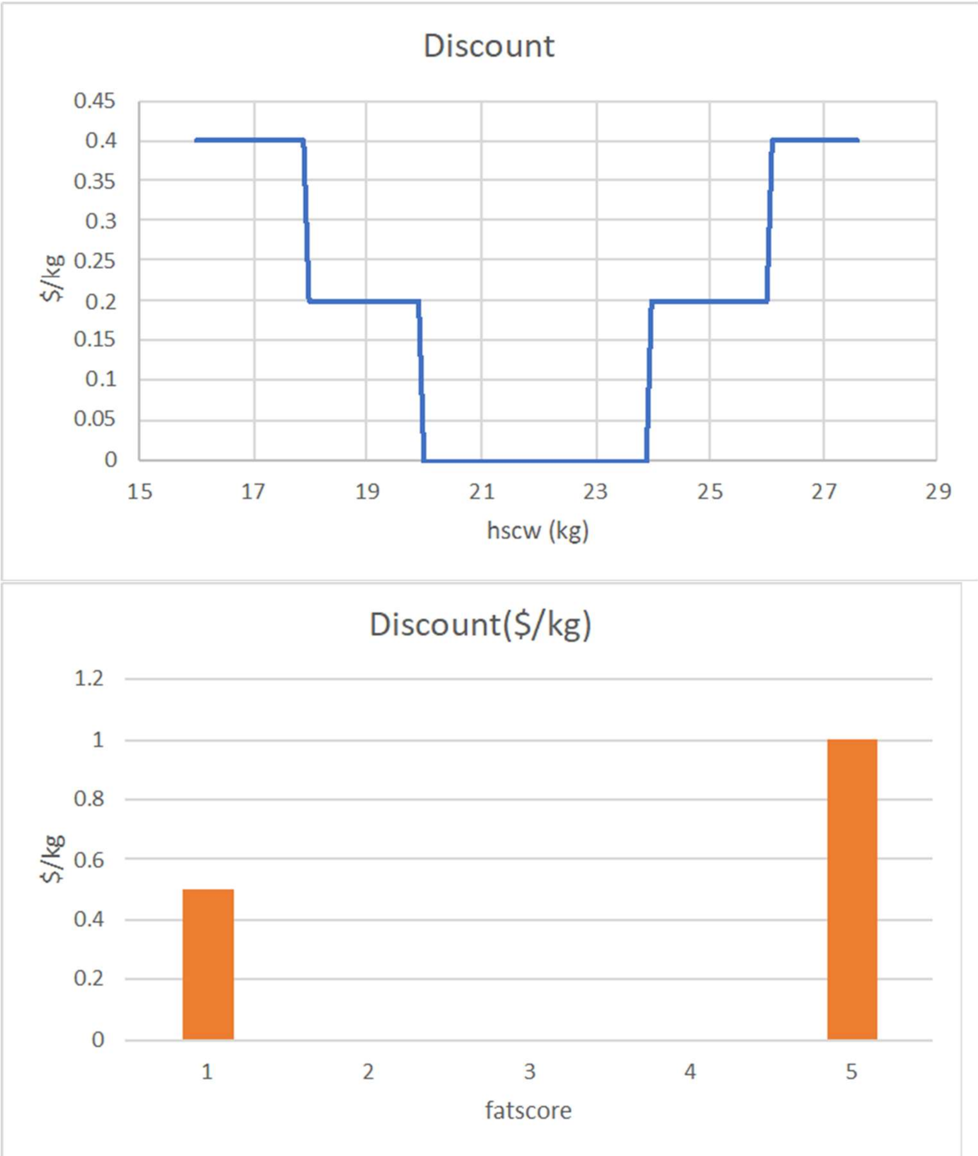

Figure 59. Grid structure for lamb value chain for which a grid control chart was constructed.

The grid control chart shows that for this producer, average losses due to non-compliance with that grid were between 2 and 6% and as expected there was a strong correlation between losses associated with carcase weight and fat-score given the relationship that those two traits have (see Figure 60).

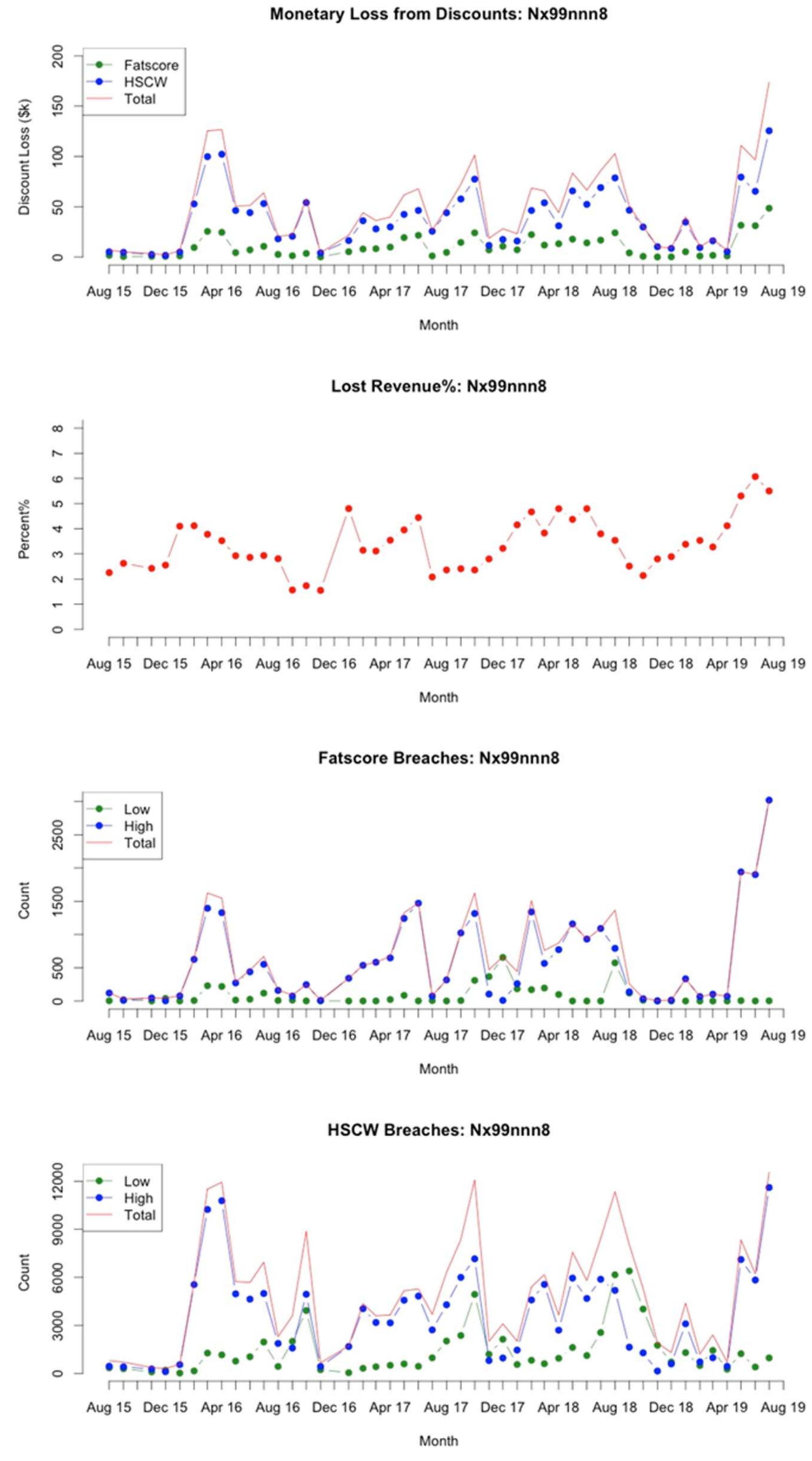

Figure 60. A grid control chart constructed for a sample lamb value grid for an individual sheep producer.

7 Conclusions

Livestock Data Link (LDL) provides the platform and framework that enables end-users (producers and processors) to interrogate data collected on carcase attributes through a range of insights related to time, traits, production characteristics and location. These insights only offer value in decision making if interpretable, acceptable and worthwhile visualisation of those traits are provided within the LDL platform.

For single traits, LDL should consider the application of tables and simple line charts or histograms. These formats provide producers will the best information on the range, scale and median of those traits and can be expressed against time and location. For two or more traits, LDL should consider the use of a combination of scatterplots and histograms.

This combination allows the user to identify and visualise correlations between traits, the variation that exists within those traits and identify any outliers both within those traits and in the combination of those traits. The format can be aggregated by location or by production type. Whilst time-based aggregation is difficult within this format, it is worth pursuing as it offers the opportunity to observe trends and make comparisons between years and seasons.

Whilst the presentation of static graphs and tables provides the end-user with some useful introductory insights, LDL should consider establishing a mechanism that enables the end-user to select and display different combinations of traits, production characteristics, locations and time. A dynamic dashboard concept has been sketched to illustrate this type of functionality. Within that dashboard, three sub-types have been developed. Importantly a time series overlay can be added to this sub-type.

The 1st sub-type produces a tabular and histogram format that is suited for an end-user considering one trait (with selection options for trait). That trait can be filtered or aggregated by production characteristics. The 2nd sub-type uses a combination of a scatterplot and histogram and is best suited for an end-user to visualise two traits that might have a correlation or relationship. Again, this sub-type can combine different combinations of traits and can aggregate based on time and production characteristics. This sub-type shows the user important features of the data including variation, range, scale and outliers both within and between the traits.

The 3rd sub-type allows the end-user to add additional overlays of thresholds or boundaries for the traits. These boundaries can be based on market specifications. This sub-type combines a table of counts and features of carcases compliant with the thresholds against those that are not and then presents that data in a scatterplot / histogram combination.

Assessing the value of insights build on historical data is difficult as often future consignments do not have the same environment or production characters. However, trends do enable an evaluation of the impact of management decisions and of choice in genetics. Calculation of the non-compliance costs are useful in determining the magnitude of investment that can be put into rectifying the non-complying factors. An alternative, but very useful graphical method of showing non-compliance is to build a grid control chart. That chart could be an automated feature of the grid wizard within LDL.

This report has identified and provided examples of a structured layered approach to the creation of aggregate or insights for carcase data. However, adoption and use of these insights will only be achieved through active training of end-users using LDL and in providing recommendations on how to adjust trait performance if the user identifies that adjustment is required to improve market compliance. A key recommendation is that the dashboard concept developed in this report should be combined with the user defined grid function. Once a user defined grid is in place the dashboard could be modified to accept the operational components of that grid.

The project team developed code that allows LDL to provide a locational benchmark of the producer's livestock performance against shire, region, state and national. However, one of the current limitations within LDL is volume of data that is available for key fields that are important aggregates or insights such as breed, bull or ram source, sex etc. It is recommended that LDL consider a R&D initiative for LDL that would allow producers to upload data into LDL on basic production characteristics such as breed and sex (sheep). This functionality would enhance the insights that are available, particularly those related to genetics, whilst protecting the core integrity of data supplied by the processor.

7.1 Recommendations

The project team have identified the following recommendations for consideration by the ISC management team. The recommendations are not ranked or listed in order of importance or magnitude of impact.

Number	Recommendation
1	That a combination of a table and line chart or histogram be used for visualisation of one trait when aggregated against a production characteristic or time series
2	That a scatterplot / histogram combination be used as standard graphical interface for both variable and factor (aggregates) when combining two traits in LDL
3	That factors are drawn in a hierarchical order from the most abundant to the least abundant to enhance visualisation when overlying a time or location time series
4	That a dynamic dashboard comprising three sub-types is considered for implementation in LDL. This would enable an end-user to customise visualisation of data for one or multiple traits
5	That the dashboard concept be expanded to incorporate grid functionality by aligning thresholds or boundaries within the grids to the 3rd sub-type of dashboard
6	That consideration is given to the construction of a grid control chart when producers enter grids as a way of graphically demonstrating the value of an insight.
7	That LDL consider the ability of producers to upload production characteristics that may not be collected at slaughter or grading, to enhance LDL insights

8 Appendix

There is no appendix for this report.