



ECOM5005 Tutorial 2

Topic 2: Organising and Presenting Data

Curtin University

School of Accounting, Economics and Finance

Aligned with Topic 2: Organising and Presenting Data



Preparation and learning objectives

Prepare before attending the tutorial. Where you cannot complete a task, record the steps you attempted and the exact point at which you became unsure. Bring the workbook and your questions to class.

After completing this tutorial, you should be able to:

- Describe one categorical variable using summary tables and charts.
- Describe one numerical variable using ordered arrays, stem-and-leaf displays, frequency distributions and graphs.
- Examine two categorical variables using contingency tables and conditional percentages.
- Compare numerical distributions using box plots and appropriate summaries.
- Visualise two numerical variables using scatter diagrams and time-series plots.
- Apply principles of honest, clear and decision-relevant visualisation.

Part 1 - Short-answer and applied problems

Question 1. Stem-and-leaf display

A statistician for a chain of department stores created the following stem-and-leaf display showing the number of pairs of glasses held at each store.

Stem	Leaves
0	9
1	3 6 6 8
2	1 2 3 5 6 9
3	0
4	1 2 3 3 5 7

Key: 4 | 1 = 41 pairs of glasses.

How many stores have no more than 30 pairs of glasses? Explain how you read the display and show the values included in your count.

Question 2. Contingency table

The contingency table below summarises a random sample of 93 students by favourite season and allergy status.

Season	Has allergies	No allergies	Total
Winter	5	9	14
Spring	6	7	13
Summer	13	23	36
Fall	14	16	30
Total	38	55	93

Which statement is true about the students surveyed? Justify your answer using the correct denominator.

- A. A student with allergies is least likely to have spring as their favourite season.
- B. A student was most likely to have summer as their favourite season, whether or not they have allergies.
- C. A student with no allergies is more likely to have spring than fall as their favourite season.
- D. A student was more likely to have no allergies, no matter what their favourite season was.



Question 3. Comparing two speed distributions

A state trooper measures the speed, in miles per hour, of every northbound and southbound car passing one highway location during a 15-minute period.

Northbound cars

60	62	62	63	63
63	64	64	64	65
65	65	65	66	66
67	68	70	83	

Southbound cars

55	56	57	57	58
60	61	61	62	63
64	65	65	67	67
68	68	68	68	71

Draw box plots of the two datasets. Then compare their centre, variability, shape and unusual observations using the plots and appropriate numerical summaries. State clearly what can and cannot be concluded from this 15-minute observation period.

Question 4. Real-life case: airport baggage delays

An airport operations team audits 930 checked bags. Each bag is classified by itinerary type and whether it arrived late at the destination carousel.

Itinerary	Not delayed	Delayed	Total
Direct flight	550	50	600
One connection	200	40	240
Two or more connections	60	30	90
Total	810	120	930

- Convert the table to percentages of the overall total.
- Calculate the delayed-bag rate within each itinerary type using row percentages.
- Among delayed bags only, calculate the percentage belonging to each itinerary type using column percentages.
- A manager states: "Direct flights are the riskiest because they produced the largest number of delayed bags." Evaluate this statement.
- Recommend one chart for communicating the operational risk to management. Explain why your chosen denominator and chart are appropriate.
- State one important limitation of drawing a causal conclusion from this table.



Question 5. Real-life case: electricity demand and temperature

A metropolitan electricity network records average daily temperature and peak demand over 14 operating days.

Day	1	2	3	4	5	6	7
Temperature (°C)	8	10	12	15	18	20	22
Demand (MW)	5200	5000	4700	4400	4150	4000	3900
Day	8	9	10	11	12	13	14
Temperature (°C)	24	26	28	30	33	36	39
Demand (MW)	3950	4100	4300	4550	4900	5250	5650

- Create a time-series plot of peak demand by day.
- Create a scatter diagram with temperature on the horizontal axis and peak demand on the vertical axis.
- Describe the pattern shown by each graph. Explain why a single straight line may provide an incomplete summary of the scatter diagram.
- A presentation truncates the vertical axis at 3,850 MW and adds three-dimensional effects. Explain why this could mislead the audience and redesign the graph using the principles of proper visualisation.
- Identify two variables not recorded here that could also affect electricity demand. Explain why the graph alone does not prove that temperature causes the observed demand pattern.



Part 2 - Data exercise: Australian unemployment trends

Purpose and files

Use the supplied workbook unemp_AUS.xlsx. It contains the worksheets UE_Data and Metadata. The data are monthly unemployment-rate series for Persons, Males and Females and are expressed as percentages. All statistical work and charts must be completed in Microsoft Excel.

Question 6. Excel analysis and dashboard

Complete the following tasks in order. Your final workbook must retain the two source worksheets unchanged and add two new worksheets named Analysis and Dashboard.

Task A - Protect and audit the data

1. Save a working copy named Tutorial 2 - Your Name.xlsx.
2. Confirm the worksheet names, data range, number of visible monthly observations, first date, last date and missing-value counts for all three unemployment series.
3. Use the Metadata worksheet to report the unit, frequency, series type and series identifiers.
4. Compare the visible observation count with the "No. Obs" entry in Metadata. Record and discuss any discrepancy rather than silently ignoring it.

Task B - Prepare an analysis sheet

Copy the complete data block from UE_Data to Analysis. Add helper columns for Year, Month and the female-minus-male unemployment-rate gap. Use formulas and fill them down to the final observation.

Task C - Produce time-series charts

Create four clear Excel line charts:

1. Persons, January 2010 to June 2020;
2. Persons, January 2000 to June 2020;
3. Persons, January 1980 to June 2020; and
4. Males and Females together, January 1980 to June 2020.

Every chart must have an informative title, labelled axes, a readable date scale, a legend where required and a sensible vertical scale. Avoid three-dimensional effects and chartjunk.

Task D - Construct summary tables

For each Persons window above, calculate the observation count, average unemployment rate, minimum and its date, maximum and its date, starting rate, ending rate and change in percentage points. Also produce a decade-style summary for the 1980s, 1990s, 2000s, 2010s and January-June 2020.

Task E - Examine gender differences

1. Plot the female-minus-male unemployment-rate gap from January 1980 to June 2020.
2. Identify the average gap, largest positive gap and date, and largest negative gap and date.
3. Calculate the change in Persons, Males and Females unemployment rates between January and June 2020.

Task F - Build a dashboard

Place the four required unemployment charts on one Dashboard worksheet. Add a title, source note and a small KPI block showing the June 2020 Persons rate, the January-June 2020 change, the 1980-2020 maximum and its date, and the 1980-2020 minimum and its date.



Task G - Write an interpretation

Write approximately 250-350 words addressing:

1. major long-run peaks and troughs;
2. how the conclusion changes when viewing 10, 20 and 40 years of data;
3. the male-female pattern and changes in the sign of the gender gap;
4. the movement between January and June 2020; and
5. limitations of the data and why a time-series graph alone cannot identify causes.

Part 3 - Pearson MyLab

1. Quiz 1 has been released on **this Tuesday (Aug-4)** at **5:00pm AWST**. There are **TEN** questions in total and you have **60 mins** to complete the quiz. Each student has **TWO** attempts. Please be aware of the due date of Quiz 1 is **Aug-11** at **5:00pm AWST**. You therefore have **SEVEN** days to complete the quiz.

If access problems remain after following the welcome-module instructions, contact the teaching team and describe the exact error message received.