

# ETF5922 & ETX2250

## Assignment Project Exam Help

Data Visualisation  
<https://tutorcs.com>  
and Analytics  
WeChat: cstutorcs

Presented by Klaus Ackermann

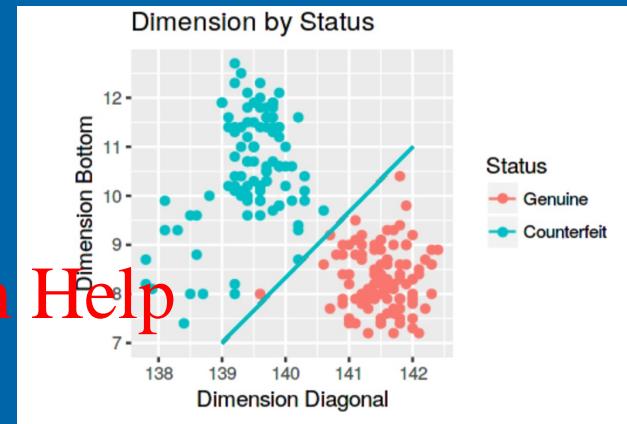
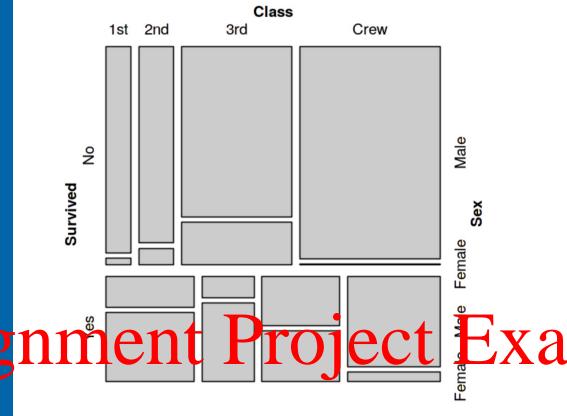
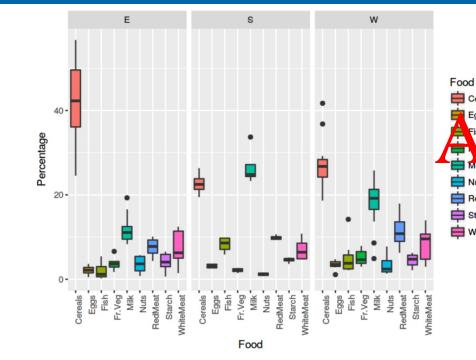


# Why Data Visualisation ?

- Gain insights into information from data by mapping it to graphical elements  
**Assignment Project Exam Help**
- Provide qualitative overview of large data sets  
**<https://tutorcs.com>**
- Search for patterns, trends, structure, irregularities, relationships among data  
**WeChat: cstutorcs**
- It helps to find interesting regions and suitable parameter for further quantitative analysis
- It is the starting point of every business analytics project

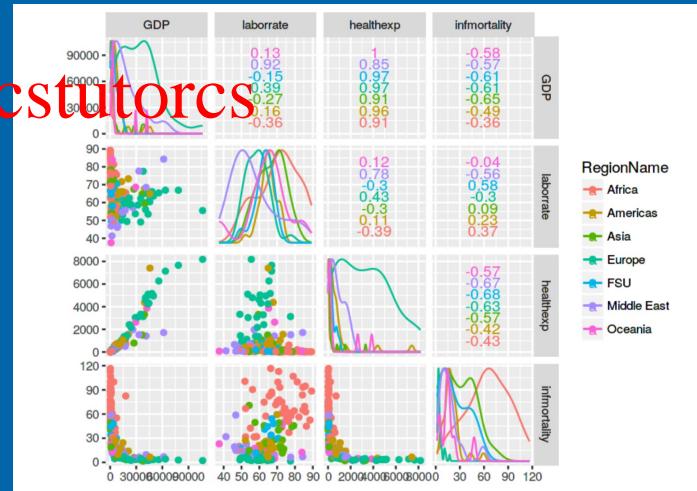
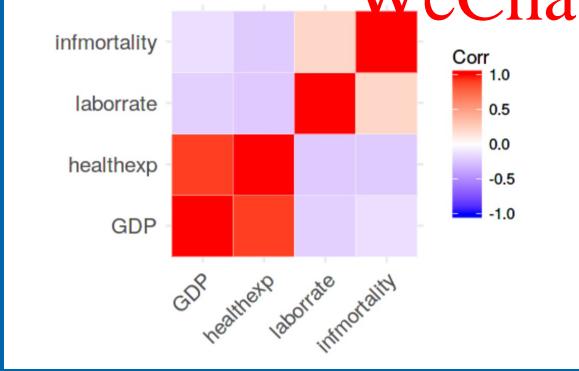


MONASH  
University



<https://tutorcs.com>

WeChat: cstutorcs



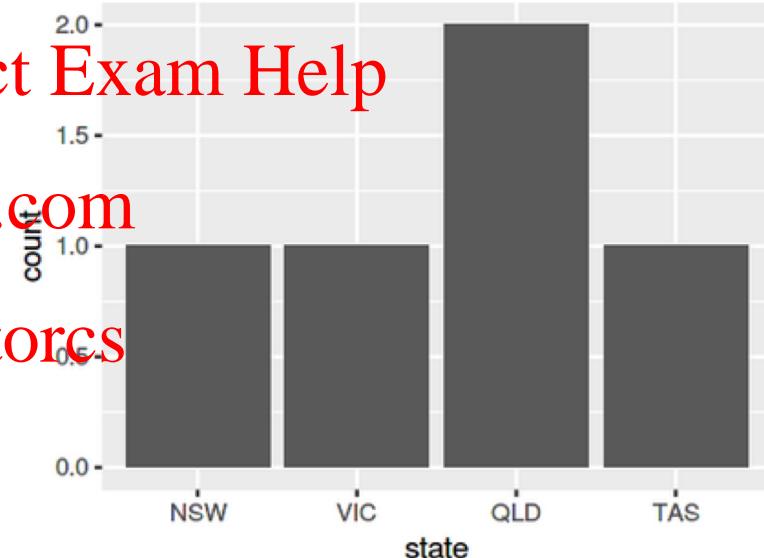
# Bar Plot

- Display of tabulated frequencies shown as bars

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



# Scatter Plot

- Provides an initial visualisation of the relationship between 2 variables

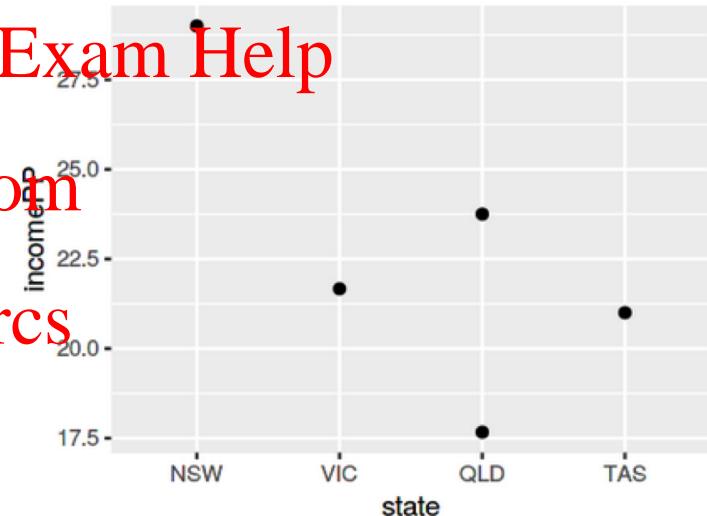
- Possible to see clusters of similar grouped points

- Each pair of values ( $x, y$ ) are treated as a pair of coordinate, then plotted on a graph

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



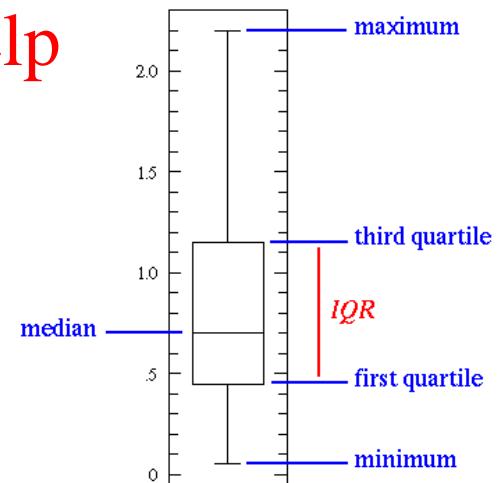
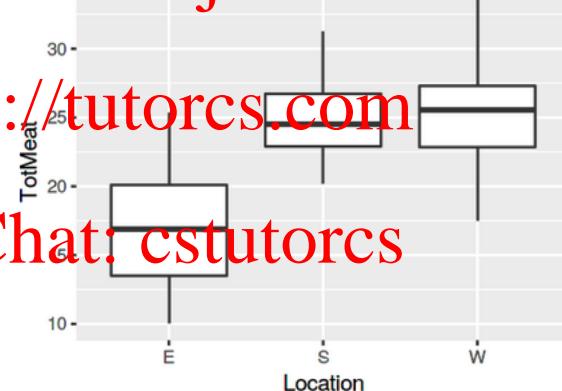
# Boxplot

- Summary of a distribution
  - Minimum, Q1, Median, Q3, Maximum
- Compare the distribution of a numerical variable across categories

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



# Steps in a Data Project

- Problem Specification

Assignment Project Exam Help

- Data Visualisation

<https://tutorcs.com>

- Data Analysis

WeChat: cstutorcs

- Evaluation & Interpretation

- And re-iterate: ideally, automate as many steps as possible

# Programming for Visualisation



- It allows us to have reproducible steps, which can be applied to many different data sets
- “common language” for a team to understand – easier to share work

**Assignment Project Exam Help**

<https://tutorcs.com>

WeChat: cstutorcs

# R- A Programming Language

- It's like learning a new language
- It has been around for a while
- It is regularly maintained and is an open source

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

```
P <- ggplot (data = protein.df,
             aes (x = RedMeat,
                  y = WhiteMeat,
                  colour = Location,
                  label = Country)) +
  geom_point () +
  geom_text (size = 3,
             check_overlap = TRUE)
```

This code allows plotting of points on a 2D plane. Further specifications are provided by adding colours to certain points to assist with visualisation

# Basic Concepts

- Assignment of a variable: `x <- 5`
- A function is a code that executes pre-defined programmed code
- A function always starts with a bracket after the name and at the end. E.g. calculate the square root: `sqrt(16)`
- Data is stored in many forms:
  - Variable
  - Factor - a variable that can only take specified values
  - Data frame, has rows and columns – similar to a table
  - Data stream: useful when you do not want to see the full data, only want to access part of the data (important when you have large data-sets)

WeChat: cstutorcs

# Grammar of Graphics



"A grammar of graphics is a tool that enables us to concisely describe the components of a graphic. Such a grammar allows us to move beyond named graphics (e.g., the ``scatterplot'') and gain insight into the deep structure that underlies statistical graphics."

Hadley Wickham.

A layered grammar of graphics.

*Journal of Computational and Graphical Statistics*, vol. 19, no. 1, pp. 3–28, 2010.

Assignment Project Exam Help  
<https://tutorcs.com>

WeChat: cstutorcs

```
ggplot(data = <DATA>) +  
<GEO4_FUNCTION>(  
  mapping = aes(<MAPPINGS>),  
  stat = <STAT>,  
  position = <POSITION>  
) +  
<COORDINATE_FUNCTION> +  
<FACET_FUNCTION>
```

# Steps for Plotting

1. How you will obtain the data  
Data frame or stream?
  - In both cases, the same commands can be used
  - Common in the course, read from a CSV, a text file where the values are separated by a comma

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

2. What do you want in the graph
  - Are you interested in all columns or just 2 at the start?

income	houseSize	adults	children	state
53	12	2	1	QLD
64	20	2	2	TAS
130	35	3	3	VIC
87	16	1	2	NSW
95	31	2	2	QLD

# Steps for Plotting

3. How do you want to represent your information
  - ie. Scatter plot vs bar graph?
  - If you only have 1 variable to work with or want to find out more about only 1 variable, start with a bar plot to get a feeling about the distribution of the data
4. Figure out what you want to include in the aesthetics of your graph
  - X, Y coordinates: how to plot it on the plane?
  - Colour: do colours create new insights?
  - Labels: Changing the axis makes it more readable and easier to present to others

WeChat: cstutorcs

# Data Manipulation

- Data will never be in the form you can use it directly

Assignment Project Exam Help

- A nice tool in R is the pipe operator: "%>%"

<https://tutorcs.com>

- It allows you to pass the information in a data frame from the left to the right to the next command, like a stream

WeChat: costutorcs

- It will make more sense in the practical part of today

state	avgIncome
NSW	29.00000
VIC	21.66667
QLD	20.70833
TAS	21.00000

# Common Issues at the Beginning

- Close the parenthesis for every function you create a bracket for (if you open a bracket, you need to close the bracket at the end)  
**<https://tutorcs.com>**
- Check parenthesis again if you are getting strange error messages  
**WeChat: cstutorcs**
- Syntax requires time to get used to, everyone struggles at the beginning

# Workshop 1



- We will go over the first section together; rest of the worksheet will be completed independently or in small groups. We will be here to assist you

**Assignment Project Exam Help**

<https://tutorcs.com>

WeChat: cstutorcs

Assignment Project Exam Help



<https://tutorcs.com>

WeChat: cstutorcs