

ETX2250/ETF5922: Data Visualization and Analytics

Introduction to visualization Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Lecturer: Lauren Kennedy

Department of Econometrics and Business Statistics

✉ lauren.kennedy1@monash.edu

CALENDAR Week 1



When have you seen
Assignment Project Exam Help
visualizations used
<https://tutorcs.com>
WeChat: cstutorcs
effectively?

A history lesson: Cholera

- In 1854 there was a breakout of the cholera disease in London killing 616 people
- At the time it was speculated that the disease was carried in the air
- A physician called John Snow was sceptical and began to collect data...

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Consequences

- The map showed the cholera was more prevalent around a water pump on Broad Street
- The pump was closed down.
- Eventually it was established that cholera is a water-borne disease.
- Data visualization saves lives!

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

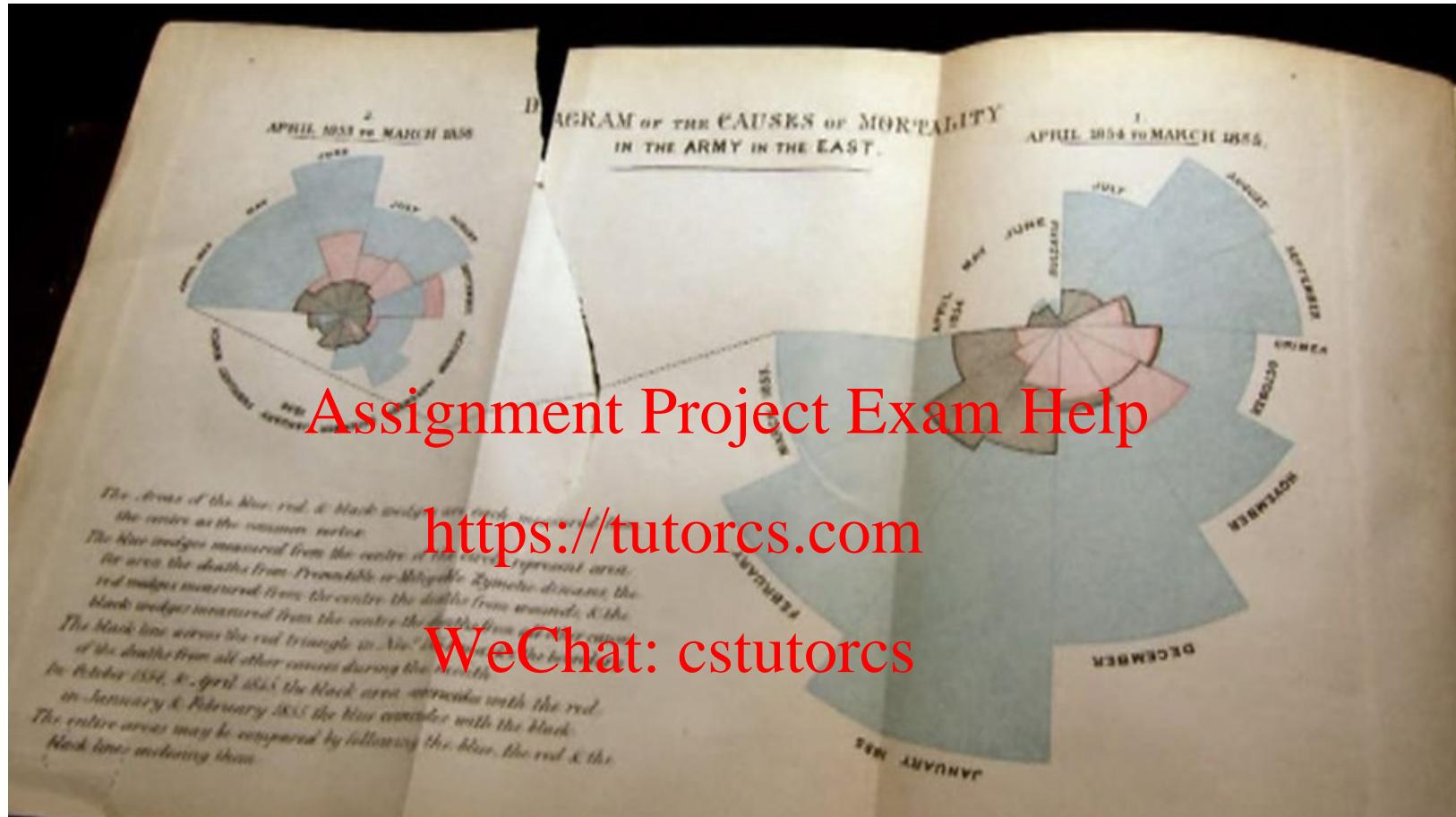
Crimean War

- At the same time Great Britain was at war against Russia in the Crimean peninsula
- Florence Nightingale is famous as a nurse who treated the wounded soldiers
- She also advocated to the British Parliament for more sanitary conditions in military hospitals.
- She knew the power of using data visualization

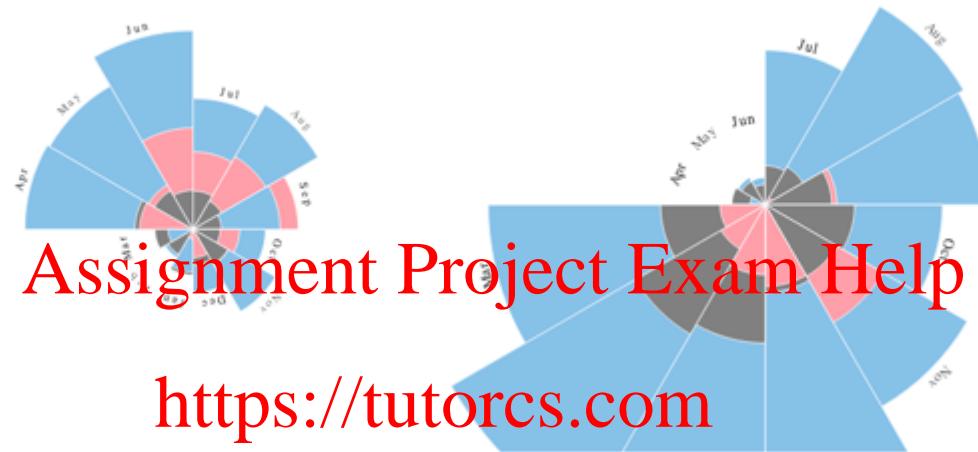
Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Nightingale's Rose chart



- Blue areas: Preventable deaths
- Red areas: Deaths from battle wounds
- Black areas: Other causes

Aftermath

- The improved sanitation at military hospitals was eventually implemented in civilian hospitals.
- Data visualisation saves lives.
- Florence Nightingale became the first female member of the Royal Statistical Society.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Napoleon

- In 1812 Napoleon thought it was a good idea to invade Russia.
- This campaign was a disaster for the French.
- Engineer Charles Joseph Minard captured the extent of this catastrophe using visualisation.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

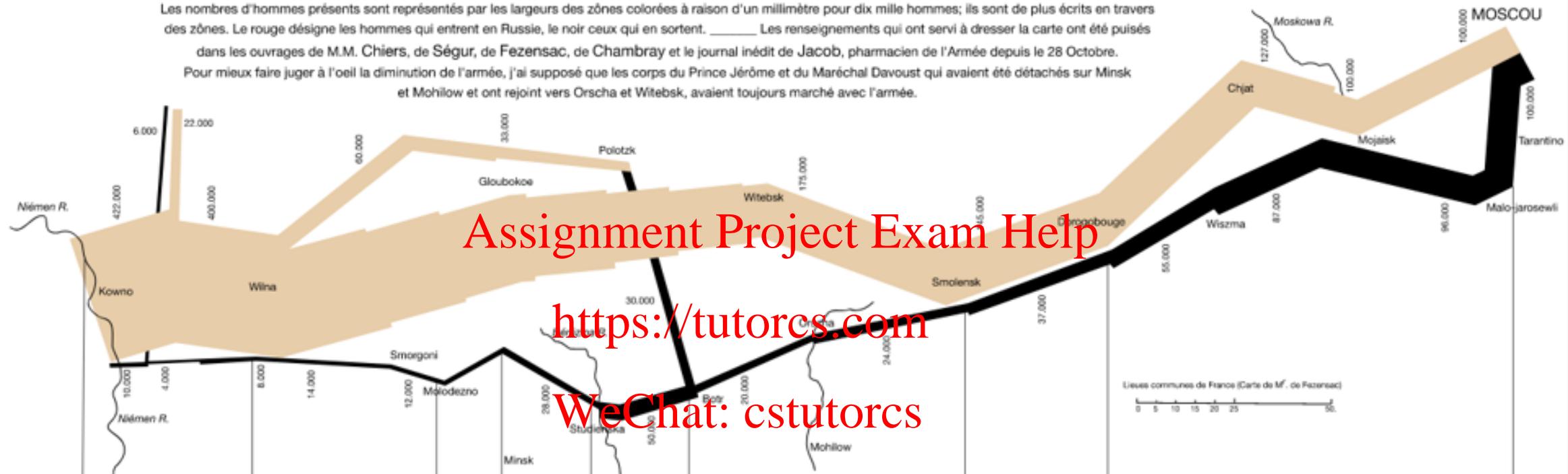
Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. _____ Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Chiens, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mohilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.



Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop
le Niémen gelé.

- 26.[°] le 7 X^{me}
- 30.[°] le 6 X^{me}
- 24.[°] le 1^{er} X^{me}
- 20.[°] le 28 9^{me}
- 11.[°]

- 21.[°] le 14 9^{me}

Début 18 8^{me}
Pluie 24 8^{me}
30 degrés

Minard's plot

This visualisation provides information on 6 variables in one chart.

- Number of troops.
- Whether troops advance or retreat.
- Temperature and time.
- Longitude and latitude.

Despite the clear message that invading Russia in winter is a bad idea, some people did not learn this lesson.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

A bit more on Minard

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Minard was so well regarded that all Ministers of Public Works in France had his work in the background of their portraits from 1850-1860

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Why data visualization?

- Gain insights into information from data by mapping it to graphical elements
- Provide qualitative overview of large data sets
- Search for patterns, trends, structure, irregularities, relationships among data
- It helps to find interesting regions and suitable parameter for further quantitative analysis.
- It is the starting point of every Business Analytics Project

[Assignment Project Exam Help
https://tutorcs.com](https://tutorcs.com)

WeChat: cstutorcs

Assignment Project Exam Help

Types of visualization

<https://tutorcs.com>

WeChat: cstutorcs

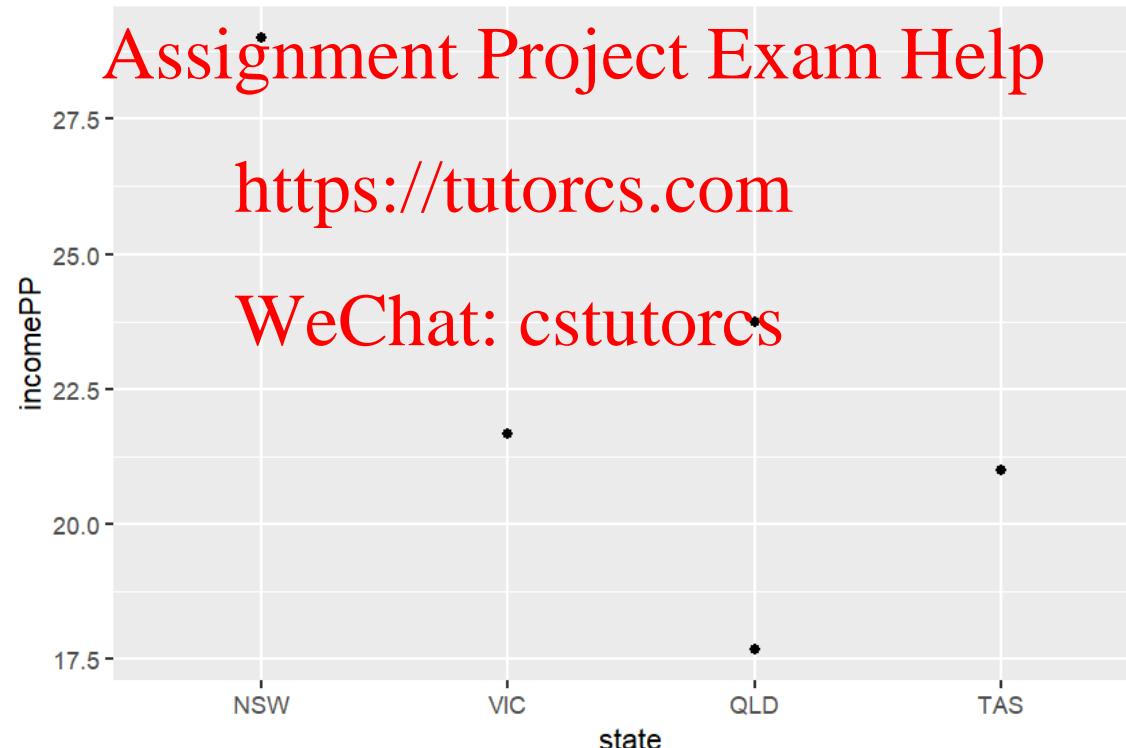
Bar plot

- Display of tabulated frequencies, shown as bars
- The height of the bar is the important factor



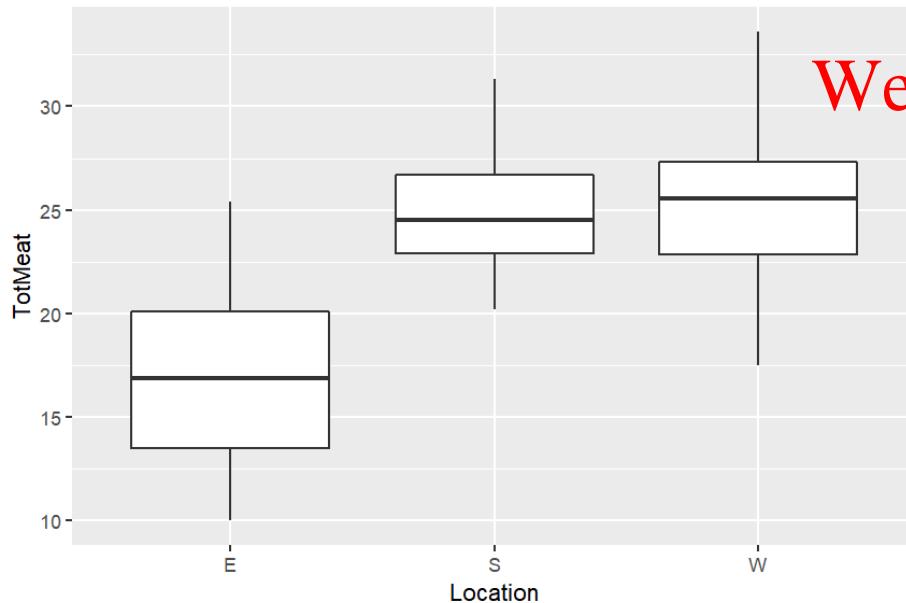
Scatter Plot

- Provides a first look of the relationship between 2 variables
- Possible to see clusters of similar grouped points
- Each pair of values is treated as a pair of coordinates and plotted



Boxplot

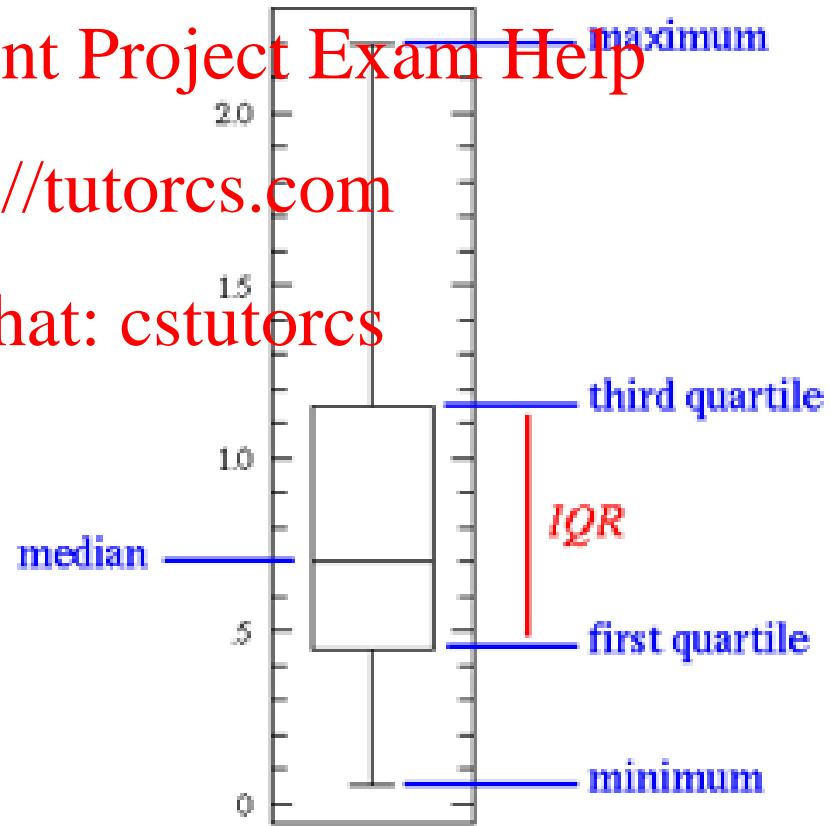
- Summary of a distribution
- Minimum, Q1, Median, Q3, Maximum
- Compare the distribution of a numerical variable across categories



Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Steps in a Data Project

- Problem Specification
- Data Visualisation
- Data Analysis
- Evaluation & Interpretation
- Communication
- And re-iterate over: Ideally, automate to ensure reproducibility at as many steps as possible

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Assignment Project Exam Help

Programming

<https://tutorcs.com>

WeChat: cstutorcs

Programming and visualization

- It allows us to have reproducible steps, which can be applied for many different data sets
- Make sure the analysis is not just point and click, you can work as a team on it on the same code
- Same goes for analysis

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

R: A programming language

- It is like learning a new language (so you need to think about learning the grammar and structure, and how to communicate well in it!)
- It has been around for a while
- It is regularly maintained and is open source
- Below you see code that plots points on a plane, colours it by country and adds the text.



<https://tutorcs.com>

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

p

Assignment Project Exam Help

WeChat: cstutorcs

Installing R

- First install R
 - Got to <https://www.r-project.org/>
 - Click "download R"
 - Select a mirror (I use the Melbourne one)
 - Install for your operating system
- Then install R studio (an IDE that we use to work with R)
 - Got to <https://www.rstudio.com/products/rstudio/download/> (you only need the free version)
 - Select download for your system
 - Follow the prompts to install
- Always interact with R using Rstudio

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Working in projects

- One challenge with working with code is having to manage your working directory and file paths (to tell the computer where things are located)
- To make this easier, we will use projects
- For the first tutorial, you will download a zipped folder from Moodle.
- Inside you will find the following files:

etx2250-etf5922-lecture-01 <https://tutorcs.com>

```
└── analysis
    └── tutorial_week01.Rmd
    └── etx2250-etf5922-lecture-01.Rproj
```

- Make sure you open by double clicking the .Rproj or by using File>Open Project in Rstudio

Glossary and important concepts

- Assignment of a variable



```
x <- 5
```

- Functions are like recipes. They take ingredients, perform a set of steps, and then return a result.



```
result <- function_name(input1 = , input2 = , ... )
```

WeChat: cstutorcs

- In R the format of a function starts with a function_name, then specify any input needed, like data

Glossary and important concepts

- Packages are collections of functions. You need to download them using the following code. Try running this command before your tutorial in the bottom left hand panel of your R studio.

```
install.packages(c("ggplot2", "tidyverse"))
```

- Once you've installed them, to use them in a specific R studio session use the **library** command

```
library(ggplot2)  
library(tidyverse)
```

<https://tutorcs.com>

WeChat: cstutorcs

Glossary and important concepts

- Data is stored in different forms in R. Two we will use are:
 - Variable
 - Data Frame, has rows and columns. Each column is a named variable
- Data has different **types** in R. Each variable must be made of the same type:
Assignment Project Exam Help
 - numeric (numbers)
 - integers (only round numbers) <https://tutorcs.com>
 - factors (a set list of possible values. Stored as labels, but also have underlying numbers)
 - characters (text data)

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.

Assignment Project Exam Help

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

<https://tutorcs.com>

WeChat: cstutorcs

Steps for plotting

Step 1. How will you obtain the data?

- Commonly in this course: read from a csv (data values separated by a common)

Step 2. What do you want in the graph?

- What information are you interested in?
Assignment Project Exam Help
- What question are you trying to answer?
- Who do you want to share what you have learned with?
<https://tutorcs.com>

WeChat: cstutorcs

Steps for plotting

Step 3. How do you want to represent your information?

- Scatter plot vs bar graph?
- How many variables are you dealing with?
- What comparisons do you want to consider?

Assignment Project Exam Help

Step 4. Figure out what you want to include in the aesthetics of your graph?

- x,y coordinates (which variable is plotted on each)
- Could you add colour to create new insights? Are the colours accessible?
- Labels - is the plot readable and easy for others to understand?

WeChat: cstutorcs

Data manipulation

- The data that you will get will never be in the form you can use it directly
- A nice tool in R is the pipe operator: "%>%"
- It allows you to pass the information in a data frame from the left to the right, to the next command. Like a stream.
- It will make more sense in the tutorial when you get to try things out

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Common issues at the beginning

- Close parenthesis for every function you open a bracket (there also needs to be a closing one)
- Check parenthesis again if you are getting strange errors
- Syntax needs time to get used too, everyone struggles at the beginning.
- Take notes of what each function does, and when you would use it. It's like learning the new words of a new language

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Assignment Project Exam Help

Live coding example

<https://tutorcs.com>

WeChat: cstutorcs



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Lecturer: Lauren Kennedy

Department of Econometrics and Business Statistics

✉ lauren.kennedy1@monash.edu

CALENDAR Week 1

