

ETX2250/ETF5922: Data Visualization and Analytics

Good and Bad 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 4



Why visualisation?

- Gain insights from data.
- Overview of large datasets.
- Search for:
 - Trends
 - Relationships
 - Irregularities
- In business data visualisation is a crucial tool to support decision making.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Tesla Motors

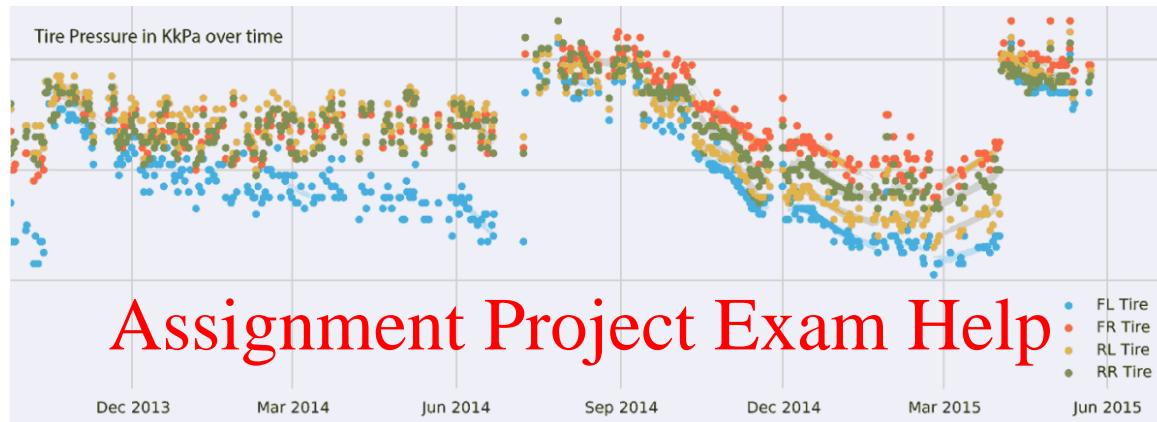
- Tesla vehicles collect a large number of data from sensors.
- The plot on the next slide shows tyre pressure over time
- This visualisation was used to
 - Check pressure when vehicles left factory,
 - See how long customers take to respond to a low pressure alert,
 - Do predictive modelling on when tyres go flat.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Tesla Motors



Assignment Project Exam Help

You can read more about the case study [here](https://tutorcs.com):

<https://tutorcs.com>

WeChat: cstutorcs

Assignment Project Exam Help

Plotting Principles

<https://tutorcs.com>

WeChat: cstutorcs

Tufte's principles

- Principles of good practice in data visualisation are outlined in *The Visual Display of Quantitative Information* by **Edward Tufte**. These include:
 - Avoid distorting what the data have to say
 - Present many numbers in a small space
 - Make large data sets coherent
 - Encourage the eye to compare different pieces of data

Assignment Project Exam Help

<https://tutorecs.com>

WeChat: cstutorcs

Bad plots

- Tufte also provides a catalog of bad plots.
- What makes these plots *bad* can be put into three categories.
 - Taste (Aesthetic)
 - Perceptual
 - Data

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Assignment Project Exam Help

Bad Taste

WeChat: cstutorcs

<https://tutorcs.com>

An ugly plot



Chartjunk

- Chartjunk is the inclusion of elements that are not necessary to communicate the information.
- The inclusion of the following can be considered chartjunk:
 - Heavy gridlines.
 - Unnecessary text.
 - Pictures within the chart.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Another example



Guidelines

Aim for a high Data Ink Ratio

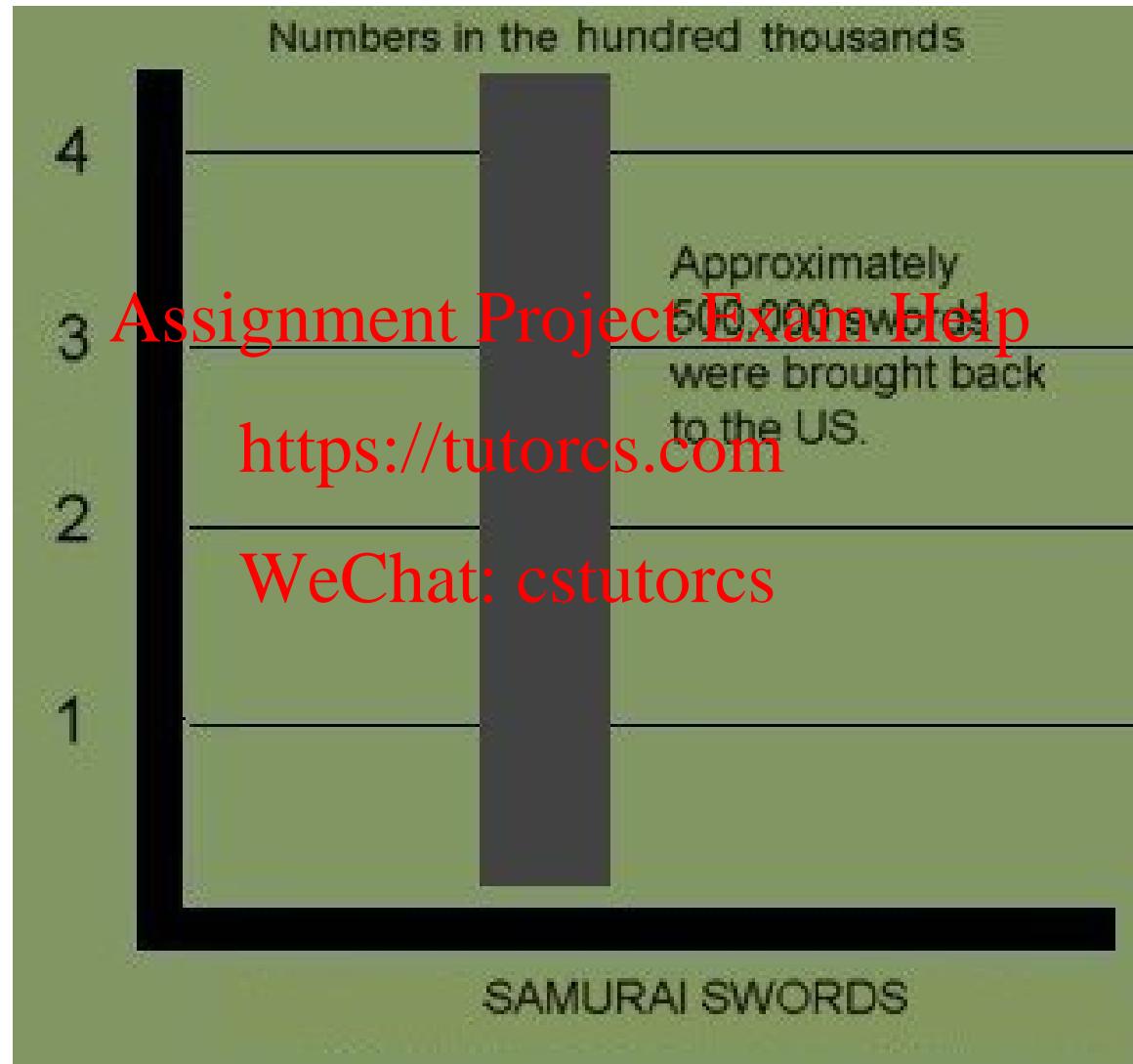
$$\text{Data Ink ratio} = \frac{\text{Ink used to display data}}{\text{Ink used in graphic}}$$

Also aim for a high Data Density [Assignment Project Exam Help](https://tutorcs.com)

$$\text{Data Density} = \frac{\text{Number of data points}}{\text{Area of graphic}}$$

If data density is small, perhaps use a table [WeChat: cstutorcs](https://tutorcs.com)

Low data density



Bad... but not misleading

- Note that although the previous plots look bad, strictly speaking they do not mislead.
- Also maximising data-ink ratio should be seen as a guideline rather than a strict rule.
- For instance the default background for `ggplot2` is arguably chartjunk
- There are good reasons for using it.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

ggplot2

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Wickham on the grey background

"We can still see the gridlines to aid in the judgement of position (Cleveland, 1993b), but they have little visual impact and we can easily "tune" them out... Finally, the grey background creates a continuous field of colour which ensures that the plot is perceived as a single visual entity"

ggplot2: Elegant Graphics for Data Analysis

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Assignment Project Exam Help

Bad Perception

WeChat: cstutorcs

<https://tutorcs.com>

What can we perceive?

- Human perception is a broad field that takes in ideas from psychology and philosophy.
- For data visualisation we can perceive:
 - Length
 - Area
 - Volume
 - Shape
 - Position
 - Color
 - Angle

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Errors of perception

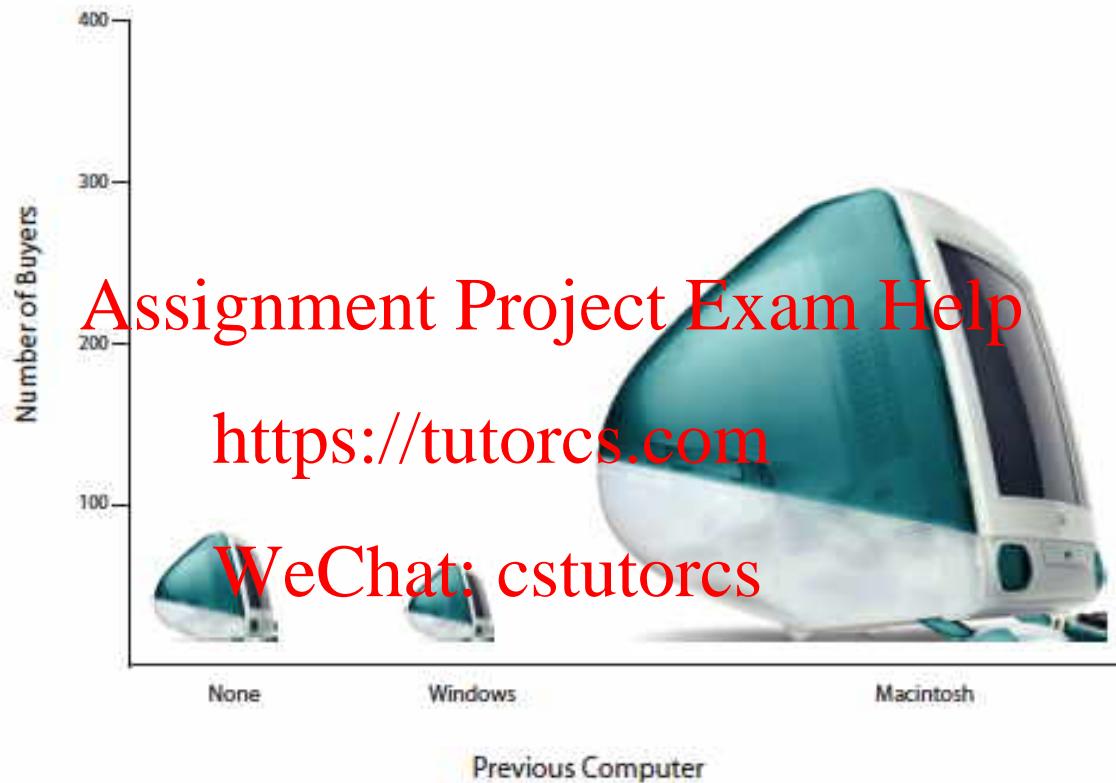
- Data visualisation is all about mapping data to things we can perceive.
- This should not be done in a way that is inaccurate or misleading.
- The following plots provide some examples of what can go wrong.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Confusing length and area



Confusing length and area

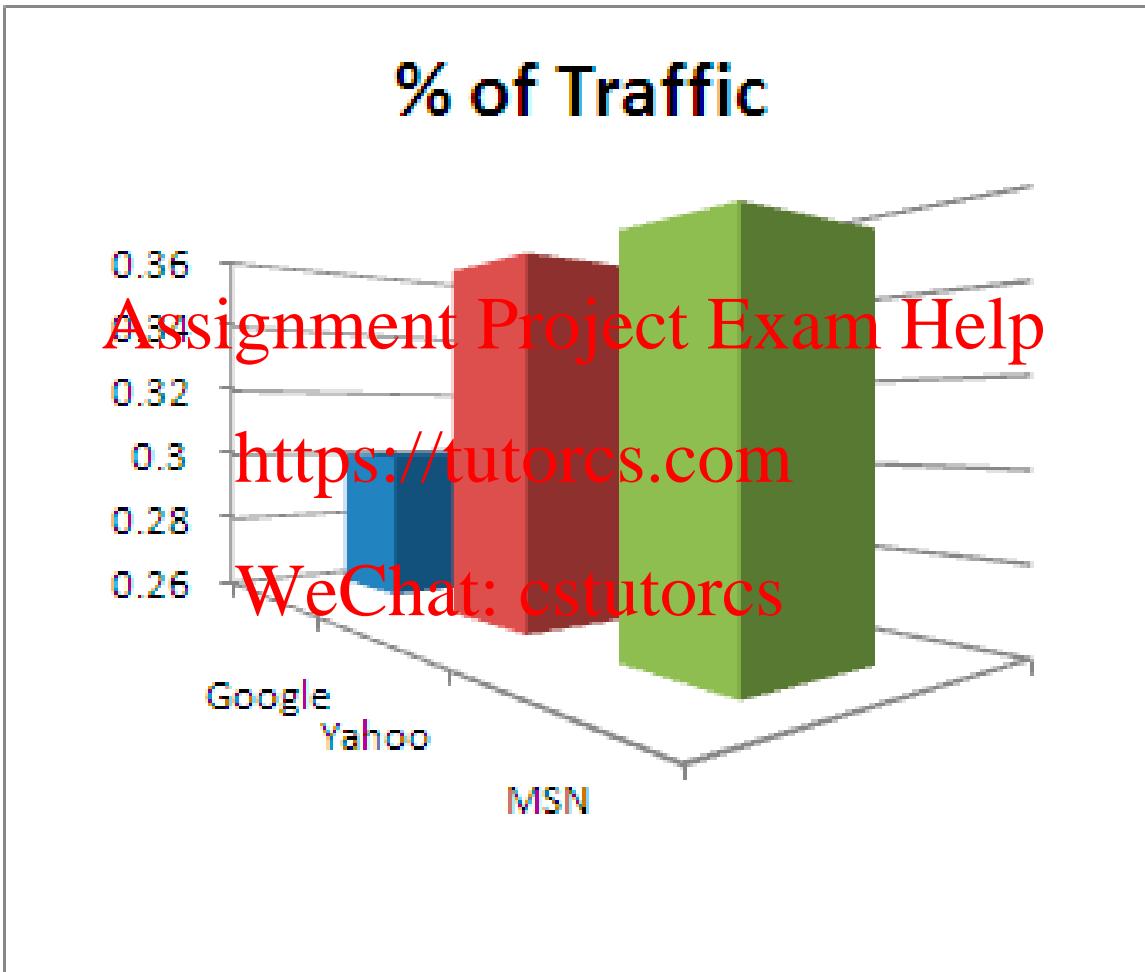
- On the previous plot the number of customers is represented by length (height of computer)
- However the *area* of the 2D pictures of computers scale up more than their *heights*.
- Also the picture leads us to imagine a 3D computer making this effect worse.
- The value for Mac is only about 3 to 4 times more than for None but we perceive the difference to be much more.

Assignment Project Exam Help

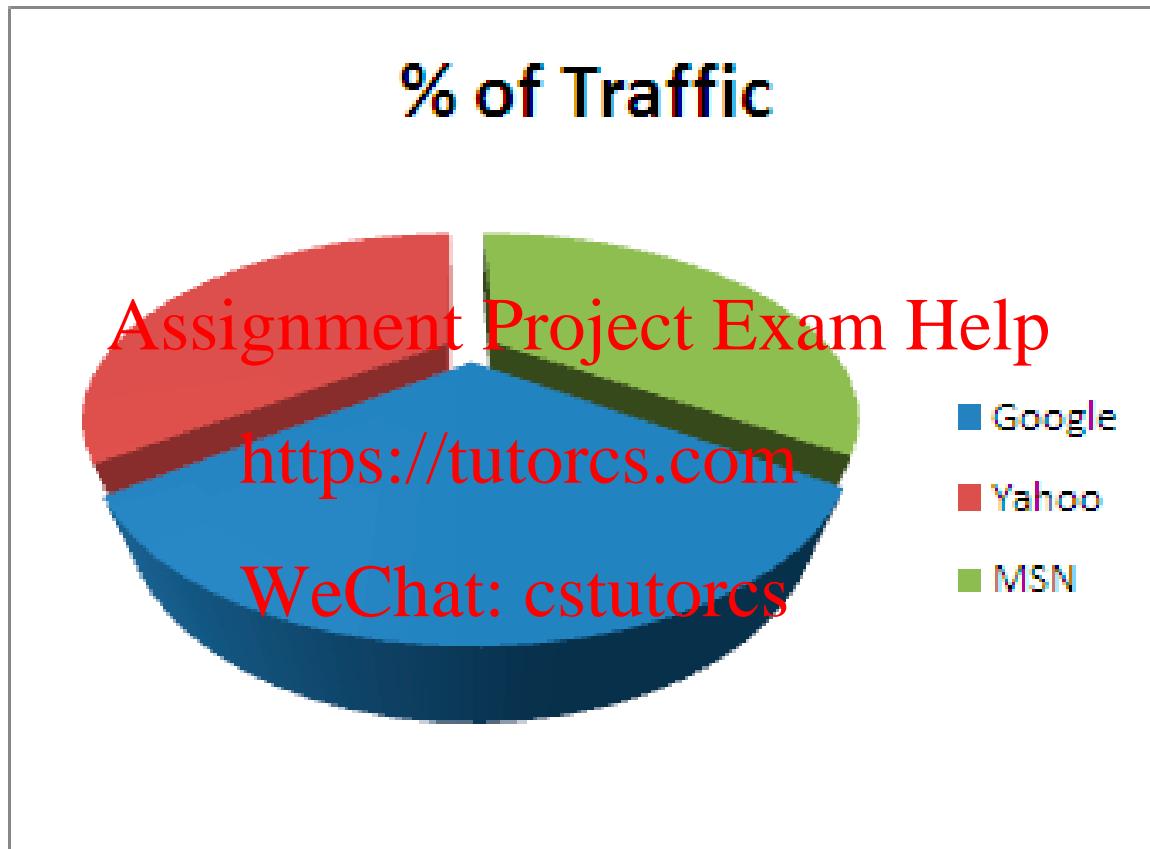
<https://tutorcs.com>

WeChat: cstutorcs

Beware 3D



Beware 3D



Beware 3D

- Difficult to line up the heights of bars with the actual values
- Closer green bar (MSN) looks bigger.
- On the pie chart rendering in 3D makes the blue segment (Google) look the biggest.
- Do not use three dimensions when two will work well.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Lie Factor

The lie factor is given by

$$\text{Lie factor} = \frac{\text{Size of effect in graph}}{\text{Size of effect in data}}$$

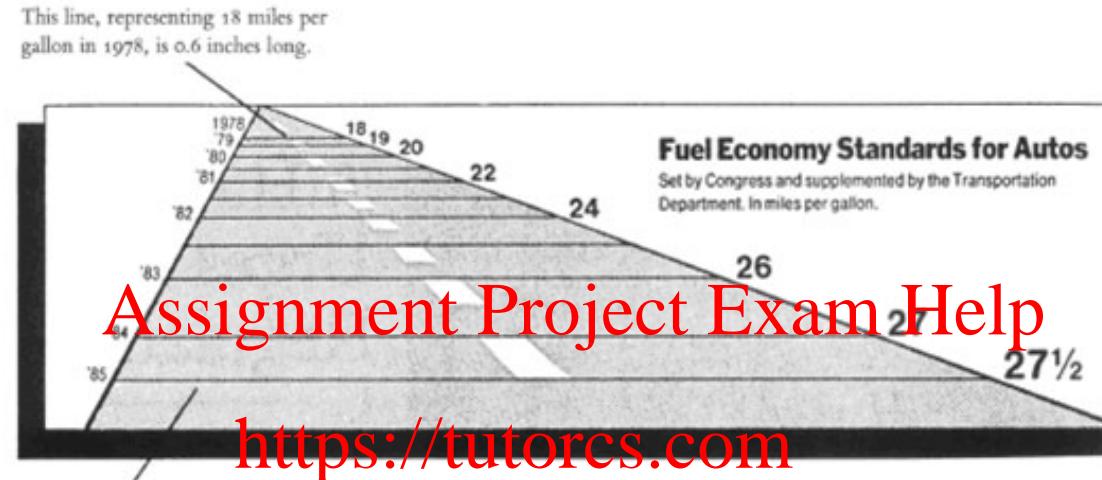
- The lie factor should be 1.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Road miles (from Tufte)



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

WeChat: cstutorcs

Effects

- The data says that mileage rose from 18 to 27.5 which is a 53% increase.
- The line on the graph increases from 0.6 inches to 5.3 inches which is a 783% increase!
- The lie factor is $783/53 \approx 14$

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Assignment Project Exam Help

Bad Data

<https://tutorcs.com>

WeChat: cstutorcs

Bad Data

- Sometimes there is nothing wrong with the plot but with the data.
- On the following slide is a plot comparing the cost of going to college in the US against the salaries of college graduates.
- Can you find problems with this graph?

Assignment Project Exam Help

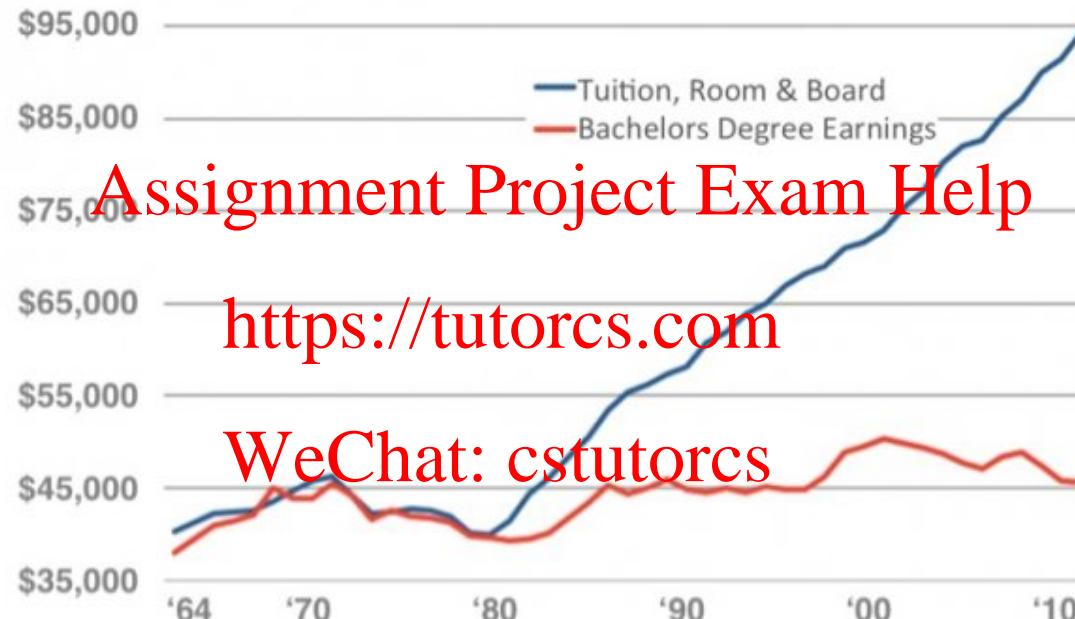
<https://tutorcs.com>

WeChat: cstutorcs

College cost

The diminishing financial return of higher education

Costs of 4-yr degree vs. earnings of 4-yr degree



Source: U.S. Census Data & NCES Table 345.

Notes: All figures have been adjusted to 2010 dollars using the Consumer Price Index from the BLS.

Problems

- There is nothing *incorrect* about this graph.
- However the message is misleading.
- The income is a yearly income while the cost of college is over four years (and only paid once).
- Also it does not show the income of people who are not college graduates.
- Think carefully about comparisons on a plot.
- Make sure your conclusions align with what is in the plot.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Assignment Project Exam Help

The x and y axis

<https://tutorcs.com>

WeChat: cstutorcs

The y-axis

- Watch this [video](#).
- Are we interested in the size of the variable rather than changes in the variable?
- Is zero a reasonable value for the variable to take?
- Are we using a bar chart?
- Answering yes to these questions means we should give more consideration to including zero on the y-axis.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Stock Prices



From this graph we conclude that Twitter stock prices increased **dramatically** on April 26.

A longer term view



Not that dramatic anymore.

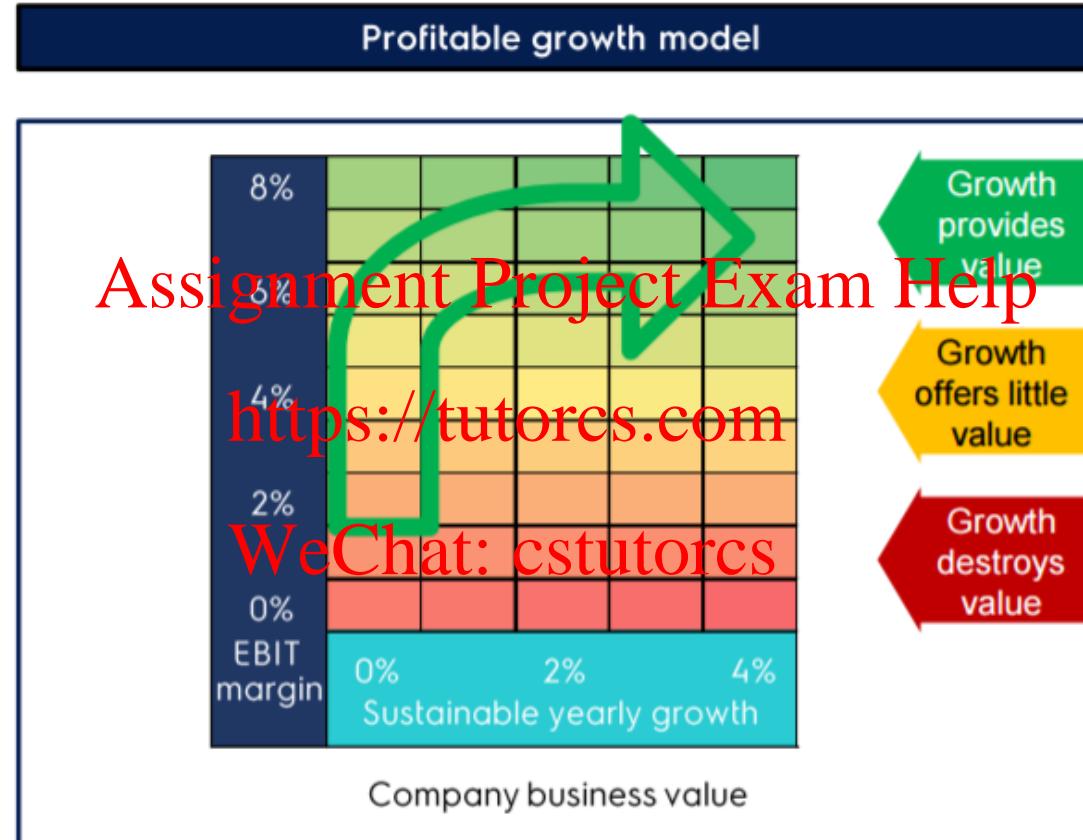
Assignment Project Exam Help

More bad plots

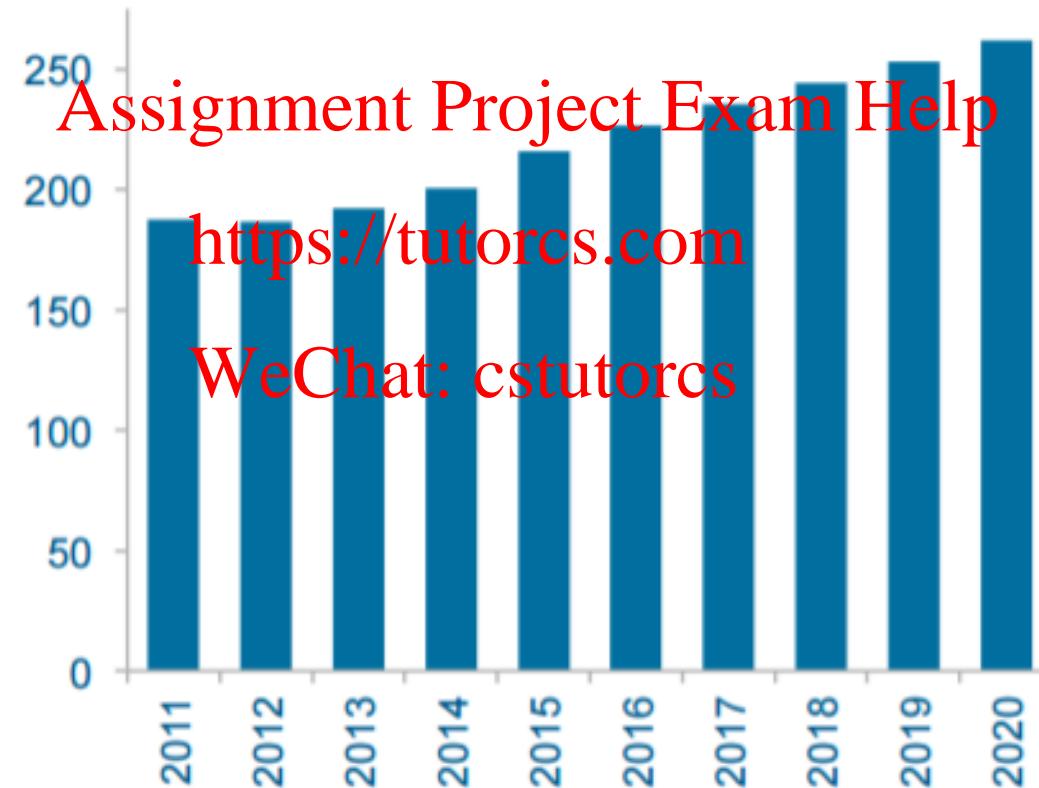
<https://tutorcs.com>

WeChat: cstutorcs

Electrolux

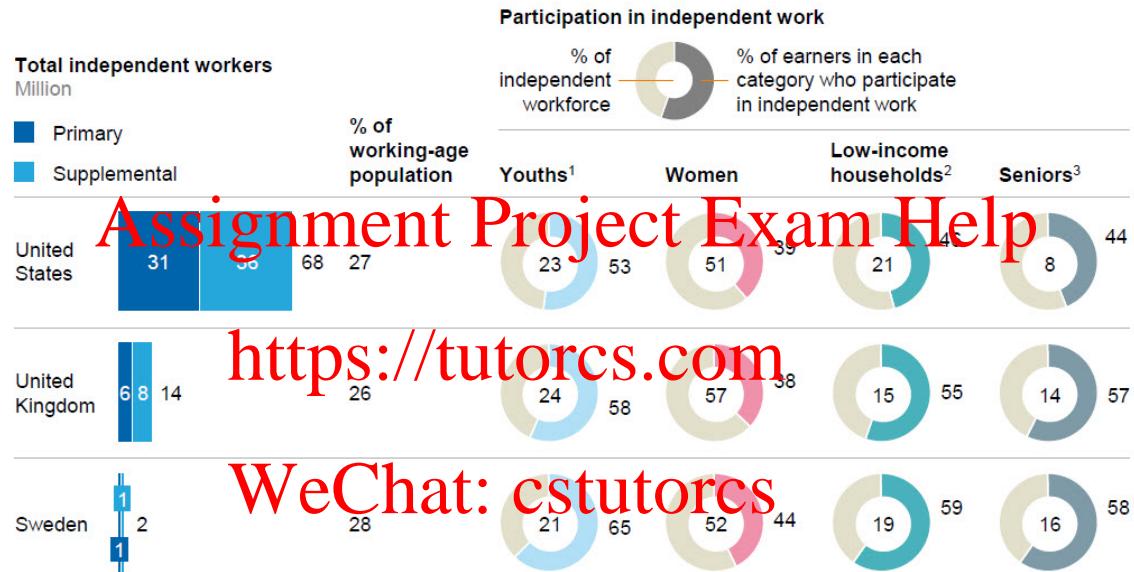


Overall European Market Sales (USD Billion)



Independent workers span all demographic groups

Responses from MGI Survey



Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

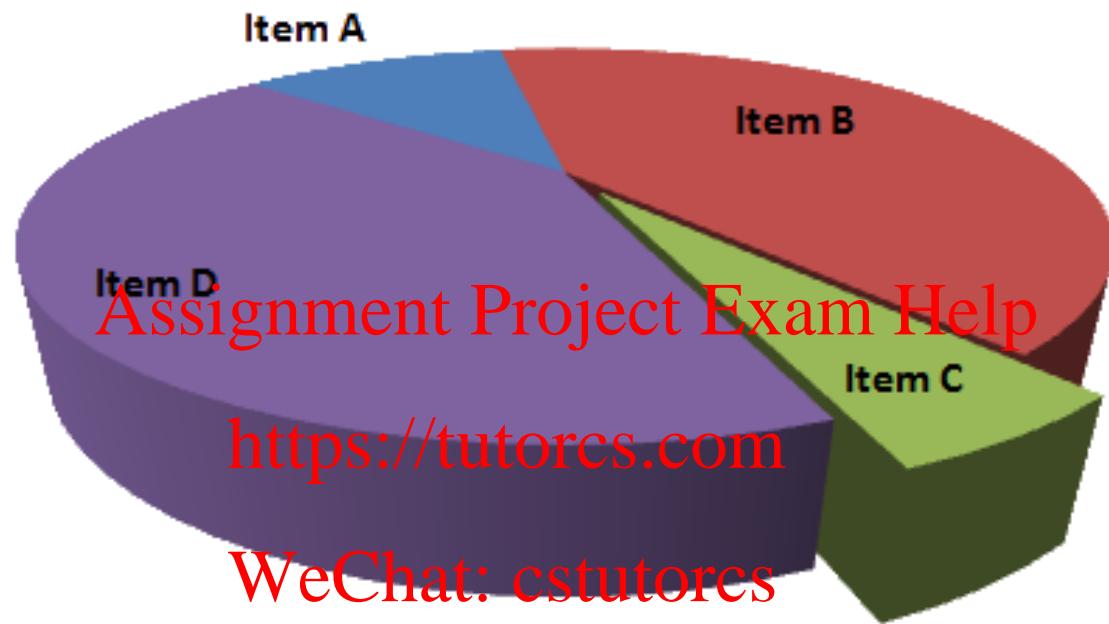
Pizza

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

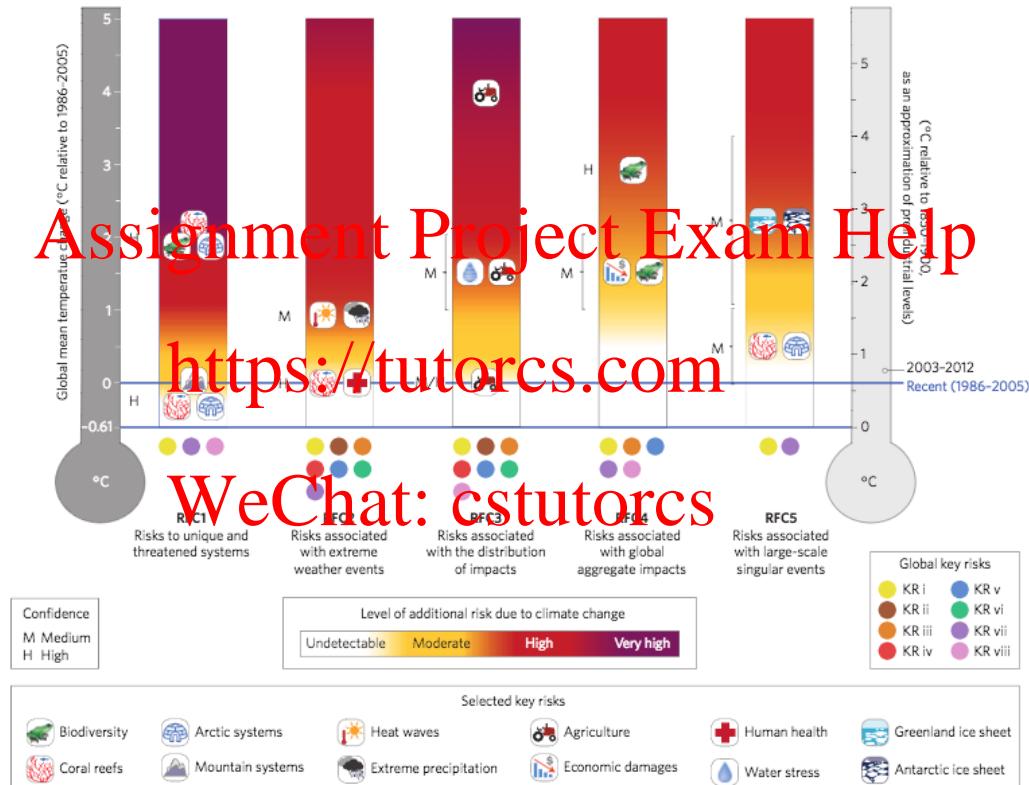
Pie chart



Climate Change

REVIEW ARTICLE

NATURE CLIMATE CHANGE DOI: 10.1038/NCLIMATE3179



Assignment Project Exam Help

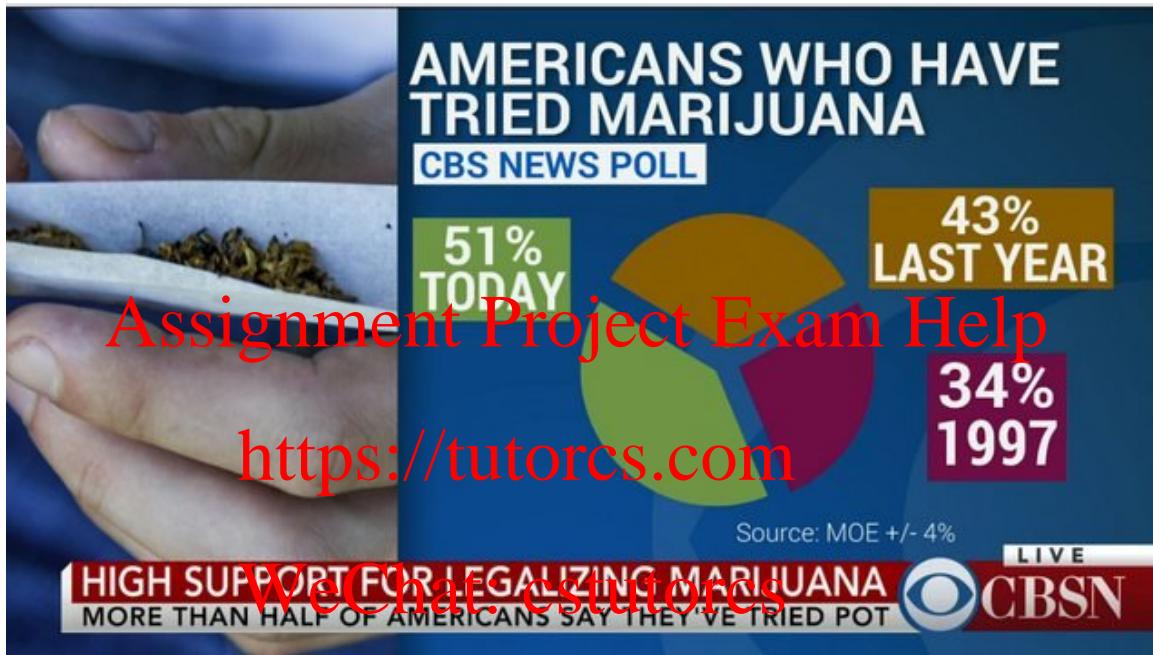
<https://tutorcs.com>

WeChat: cstutorcs

Season



Narcotics



Summary

- Graphs can be misleading
- The default options in ggplot are chosen to protect the user from errors of taste and errors of perception.
- Nothing protects you from using bad or misleading data...
- ...except for your own common sense

Assignment Project Exam Help

<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 4

