

ETX2250/ETF5922: Data Visualization and Analytics

Advanced 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 5



Visualising many variables

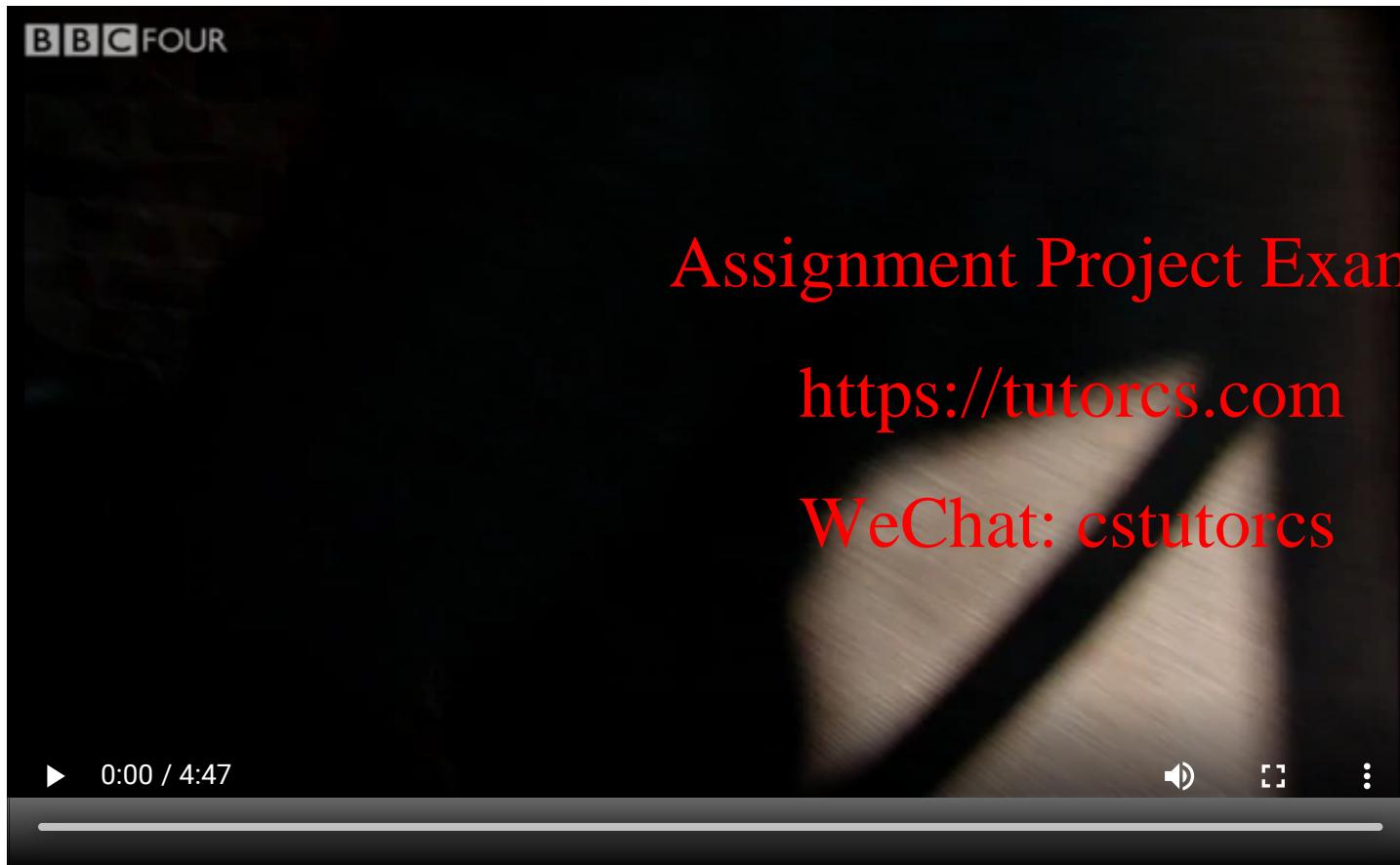
- We can do more than visualise variables spatially
 - Colour
 - Size
 - Label
 - Facets

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

An example



Mpg data

- The variable `cty` measures fuel efficiency of different cars in the city, while `displ` measures the size of the engine.
- These are negatively correlated.
- We can also see how the non-metric variable `drv` interacts with these variables using the `col` (colour) aesthetic.

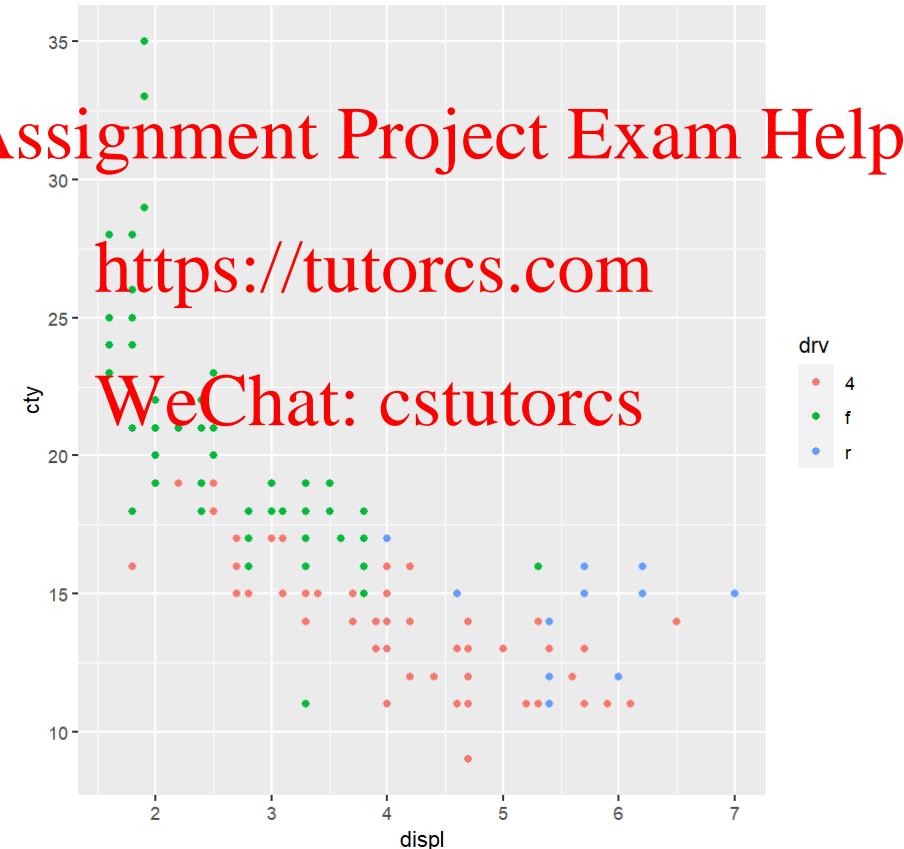
Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Using color

```
ggplot(data = mpg, mapping =  
       aes(x=displ,y=cty, col=drv))+geom_point()
```



<https://tutorcs.com>

WeChat: cstutorcs

Aes v geom

- Note that unlike previous lectures, color is being used here to display information about a variable in the dataset.
- Therefore instead of specifying color in the geom, it has to be specified in the `aes` function.
- Remember the `aes` function maps data to something we can perceive.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Text labels

- Another option is to plot text rather than points
 - This is in fact a different geom called `geom_text`
- A variable can be mapped to the actual text that appears
 - The aesthetic is `label`

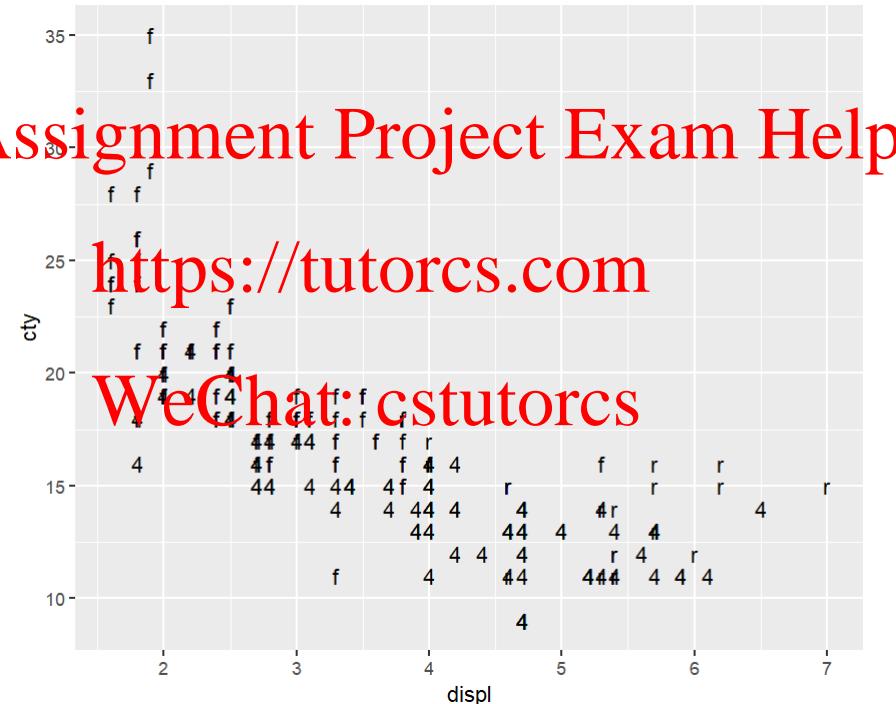
Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

With text

```
ggplot(data = mpg, mapping =  
       aes(x=displ,y=cty, label=drv))+geom_text()
```



The bubble chart

- To add a fourth variable we can manipulate the size of the points.
- This is known as a bubble chart.
- The aesthetic in question is `size`
- The following plot maps the number of cylinders to the size of points.

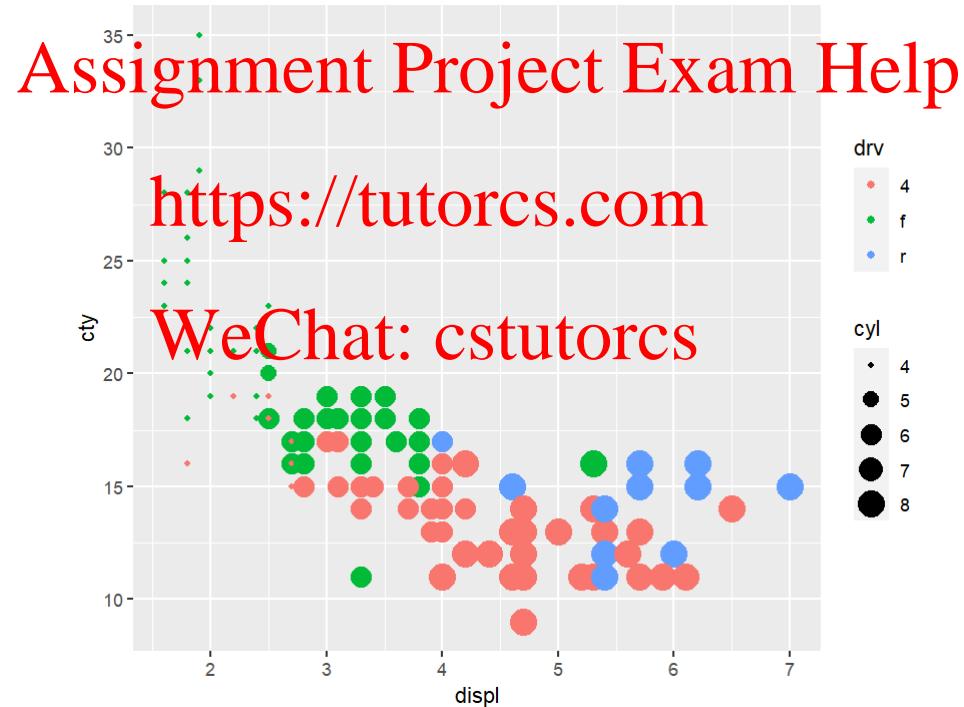
Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Bubble plot

```
ggplot(data = mpg, mapping =  
       aes(x=displ,y=cty, col=drv, size=cyl))+  
  geom_point()
```



Assignment Project Exam Help

All about colourmaps

<https://tutorcs.com>

WeChat: cstutorcs

Color scales

- Suppose we are mapping metric or ordinal data to a colormap. The colormap should be
 - Sequential
 - Perceptually uniform
 - Work when printed in black and white
 - Accessible to colorblind people
 - Colorful and pretty
- The *viridis* colormap was developed with this in mind

Assignment Project Exam Help

<https://tutorcs.com>

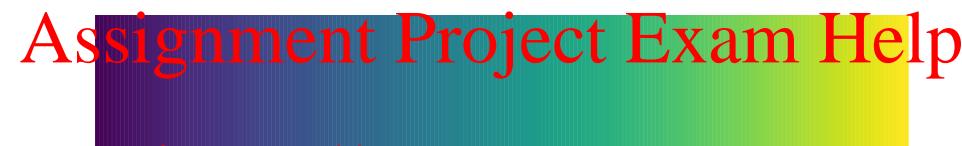
WeChat: cstutorcs

Jet v Viridis

A popular palette is jet.



A better palette (by the above criteria) is viridis



<https://tutorcs.com>

WeChat: cstutorcs

Problems with jet

- Colors close to one another should be similar.
- On jet, in some parts the color changes dramatically over a small range.
- Also colorblind people (about 8% of the population) can have difficulty with the red colors in jet.
- For more on this see [this talk](#) by the creators of viridis.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Jet Colormap

```
knitr::include_graphics('images/lecture-05/mona-lisa-rainbow.png', dpi = 100)
```

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Viridis colormap

```
knitr::include_graphics('images/lecture-05/mona-lisa-gradient.png', dpi = 100)
```

Assignment Project Exam Help

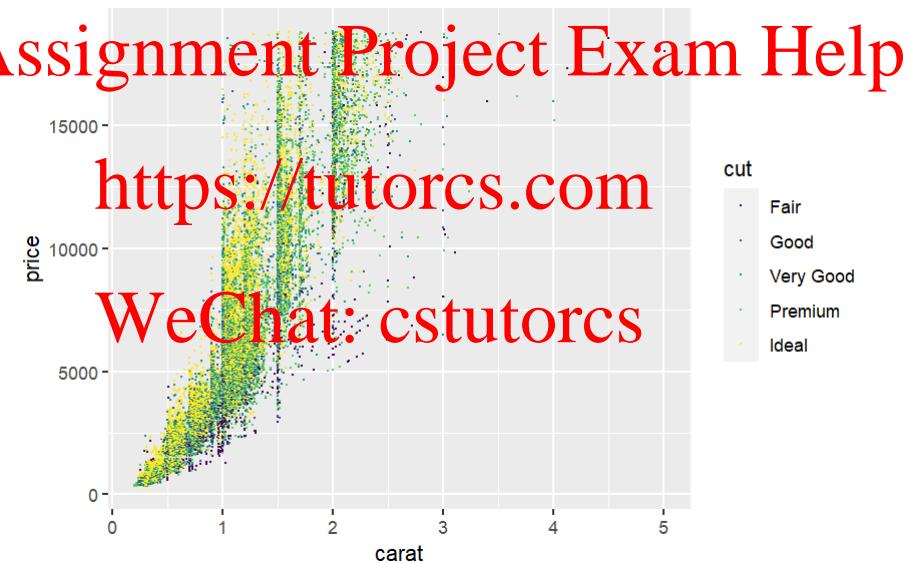
<https://tutorcs.com>

WeChat: cstutorcs

In ggplot2

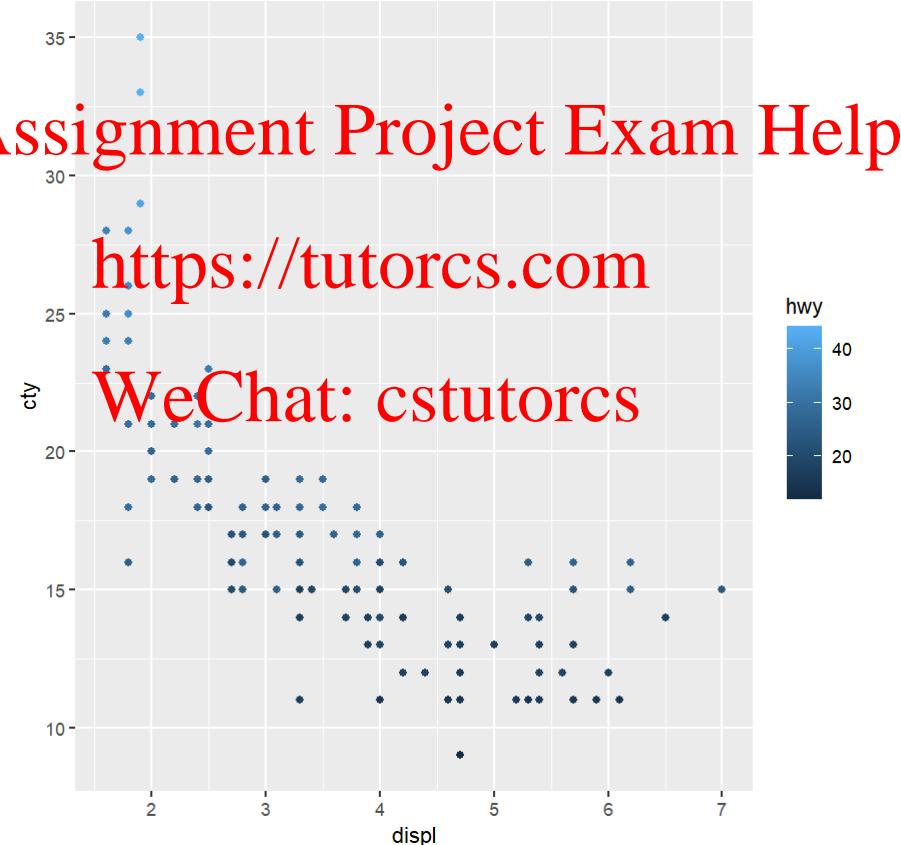
Ordered factors now use viridis by default.

```
ggplot(diamonds, aes(y=price, x=carat, col=cut))+  
  geom_point(size=0.2)
```



Continuous color

```
ggplot(data = mpg, mapping =  
       aes(x=displ,y=cty, col=hwy))+geom_point()
```



Continuous color

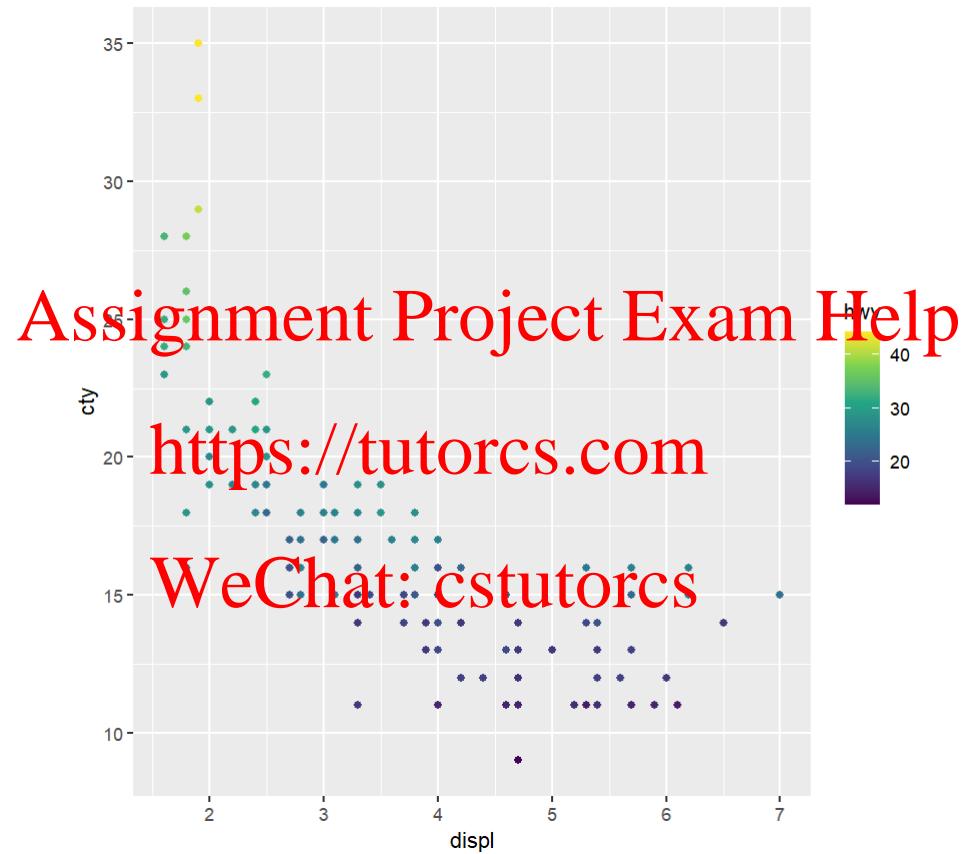
- To use viridis for a continuous variable simply add `scale_color_viridis_c()`.
- Scale is another element of the grammar of graphics.

```
ggplot(data = mpg, mapping =  
        aes(x=displ,y=cty, col=hwy))+  
  geom_point()+scale_color_viridis_c()
```

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

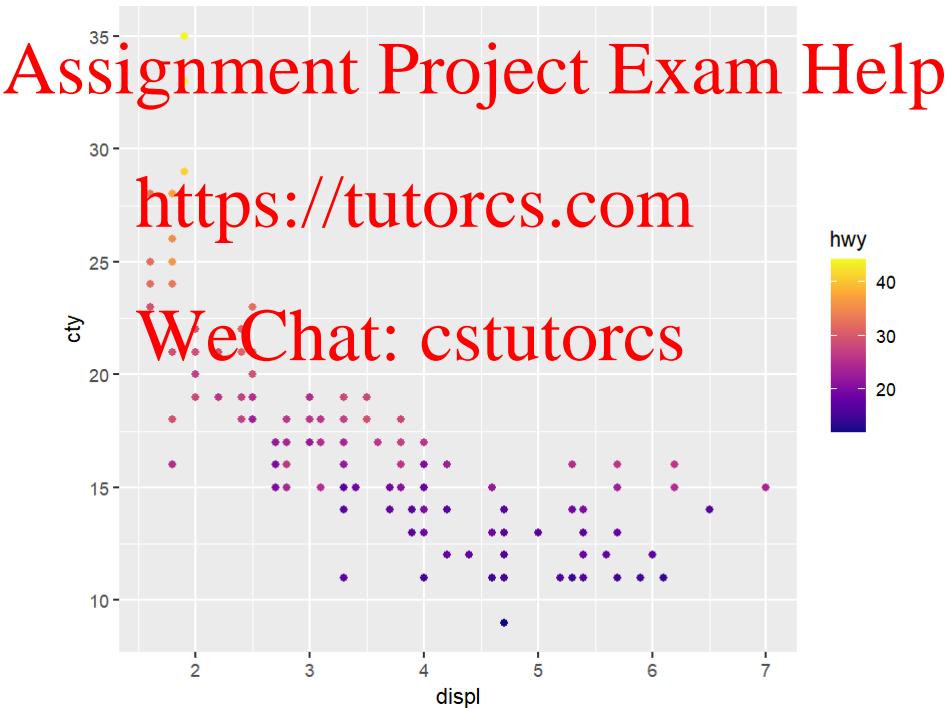
WeChat: cstutorcs

Viridis



Variations on Viridis

```
ggplot(data = mpg, mapping =  
       aes(x=displ,y=cty, col=hwy))+  
  geom_point()+scale_color_viridis_c(option = 'C')
```



Caution

- There are some situations where viridis may not be ideal.
 - Nominal variables
 - Divergent scales
- Divergent scales can be used when there is a natural middle point for the data (usually zero).
- For when plotting budget or trade balances using color, red can be used to show deficit and blue can be used to show surplus.

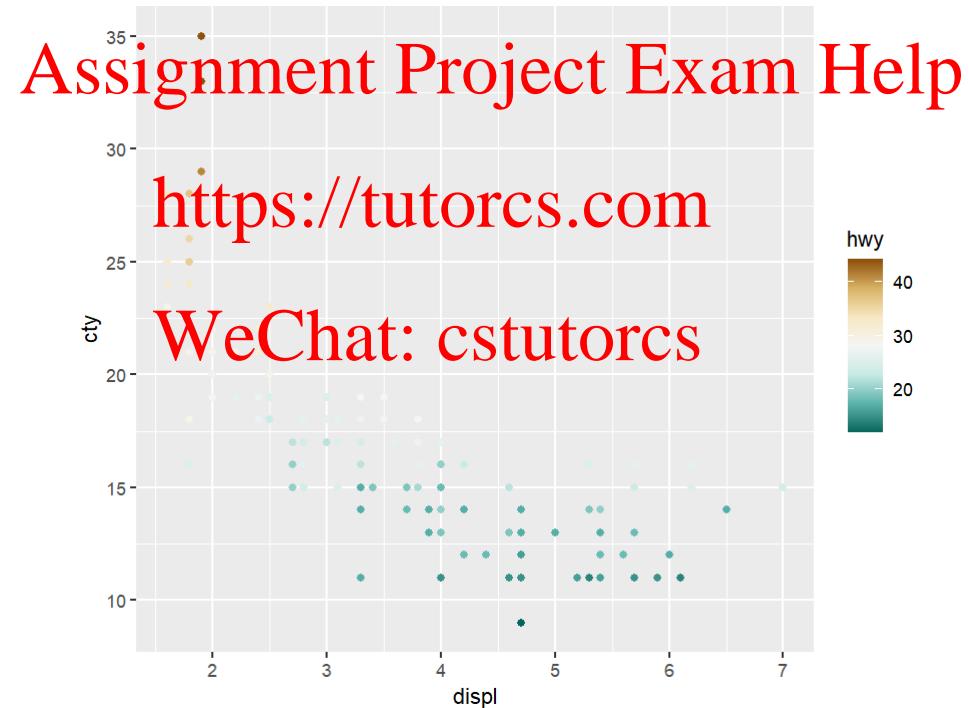
Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Divergent Scale

```
ggplot(data = mpg, mapping =  
       aes(x=displ,y=cty, col=hwy))+  
  geom_point()+scale_color_distiller(type = 'div')
```



Assignment Project Exam Help

Facetting
<https://tutorcs.com>

WeChat: cstutorcs

Facetting

- Sometimes we cannot display everything on a single plot
- In this case facetting can be used to construct multiple plots
- For the next example we look at the [Palmer Penguins](#) dataset. You can use the tidyTuesdayR package to load this data set in

Assignment Project Exam Help

```
library(tidyTuesdayR)
penguins <- tt_load(2020, week = 1) $penguins
## WeChat: cstutorcs
##     Downloading file 1 of 2: `penguins.csv'
##     Downloading file 2 of 2: `penguins_raw.csv'
```

Code for facetting

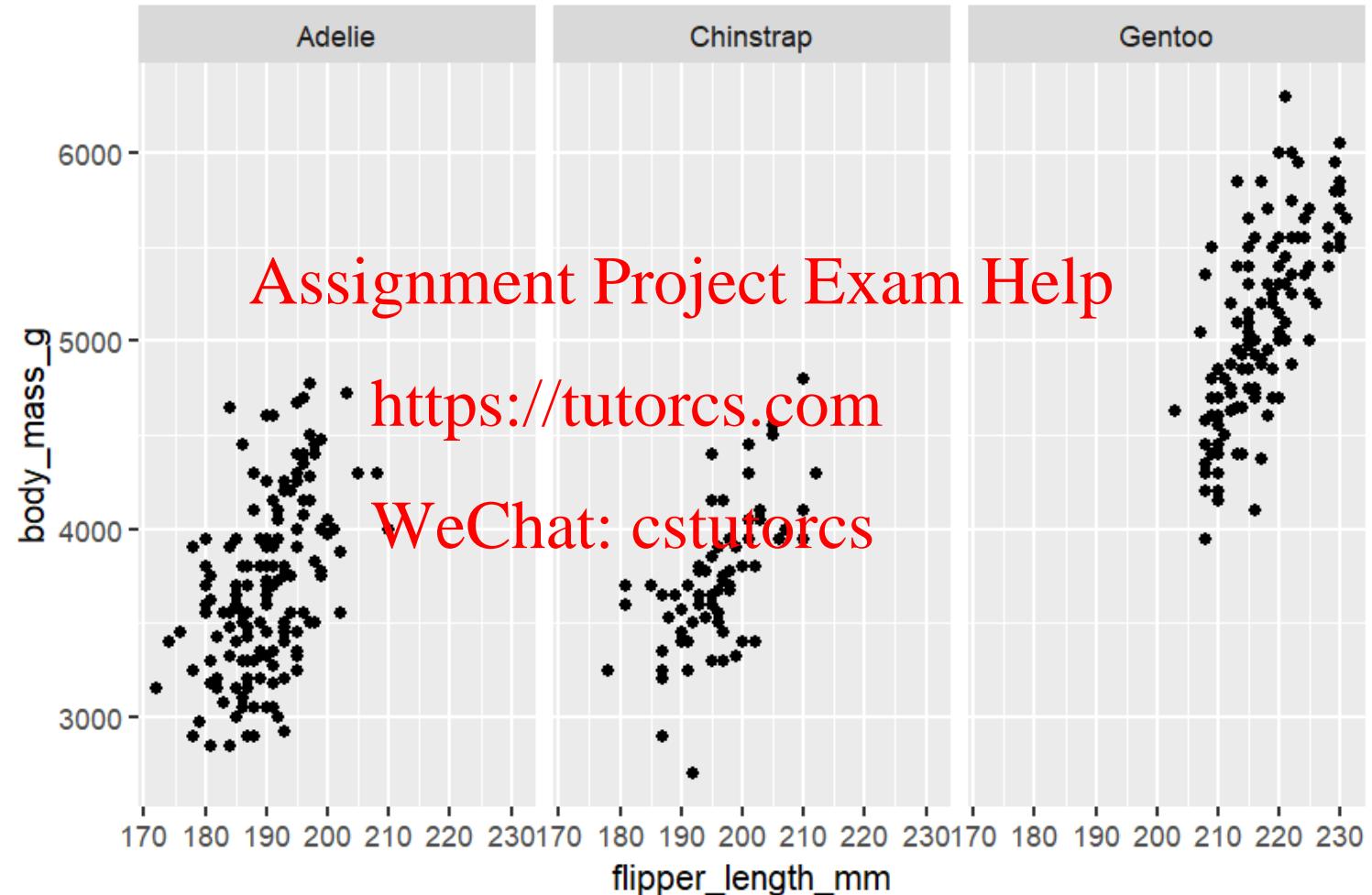
```
ggplot(data = penguins,  
       mapping = aes(x=flipper_length_mm, y=body_mass_g))+  
  geom_point() +  
  facet_wrap(~species)
```

Note the tilde (~) in `~species` Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Palmer Penguins



Scales

- Sometimes if the scales on the y axis are very different, we can't see differences between the facets.
- The option `scales` in the `facet_wrap` function allows each plot to have its own scale.
- Use this with caution! This is NOT appropriate in this case. Can you see why?

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Free scales

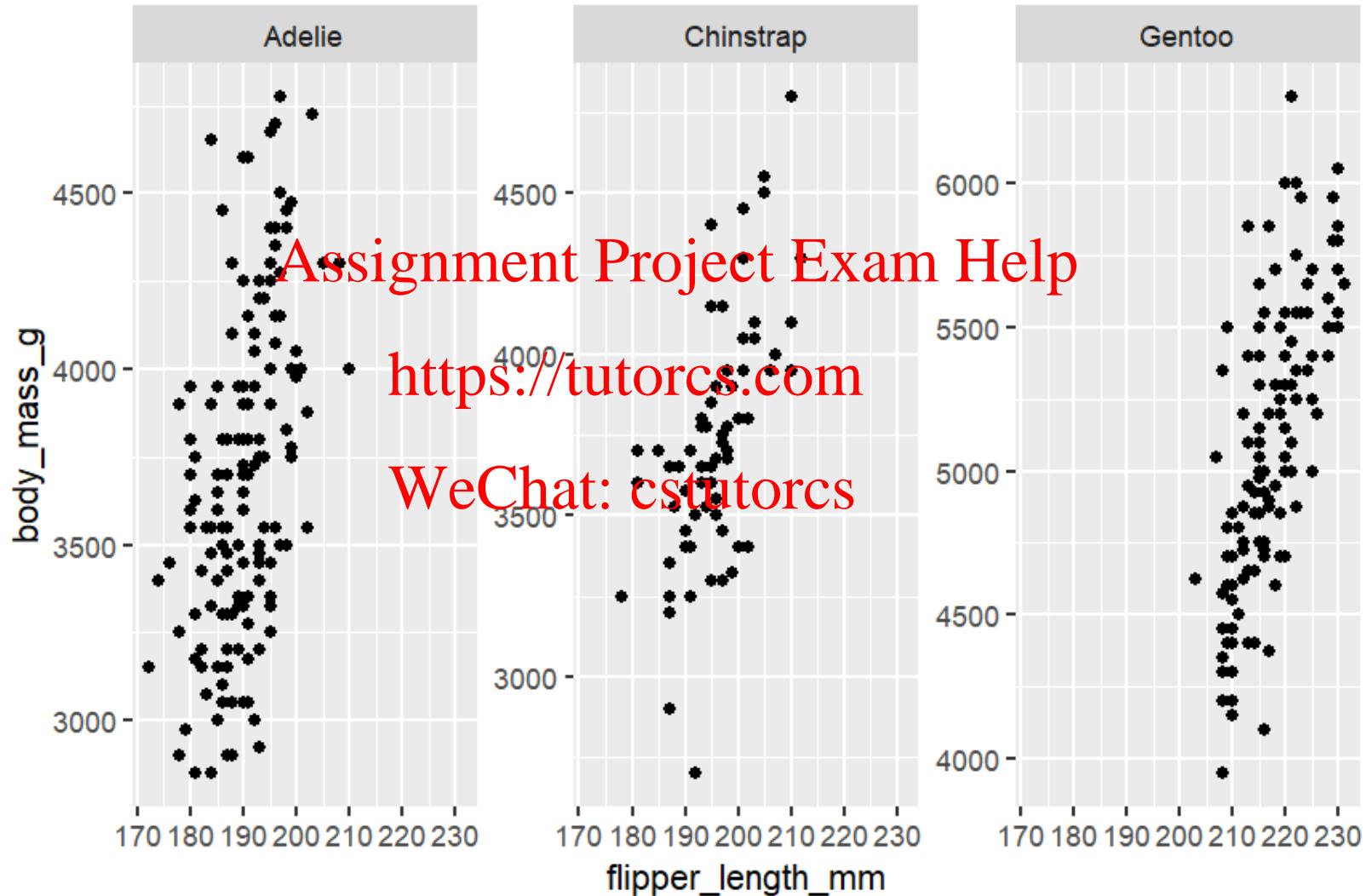
```
ggplot(data = penguins,  
       mapping = aes(x=flipper_length_mm, y=body_mass_g))+  
  geom_point() +  
  facet_wrap(~species, scales = 'free_y')
```

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Palmer Penguins



Change number of columns

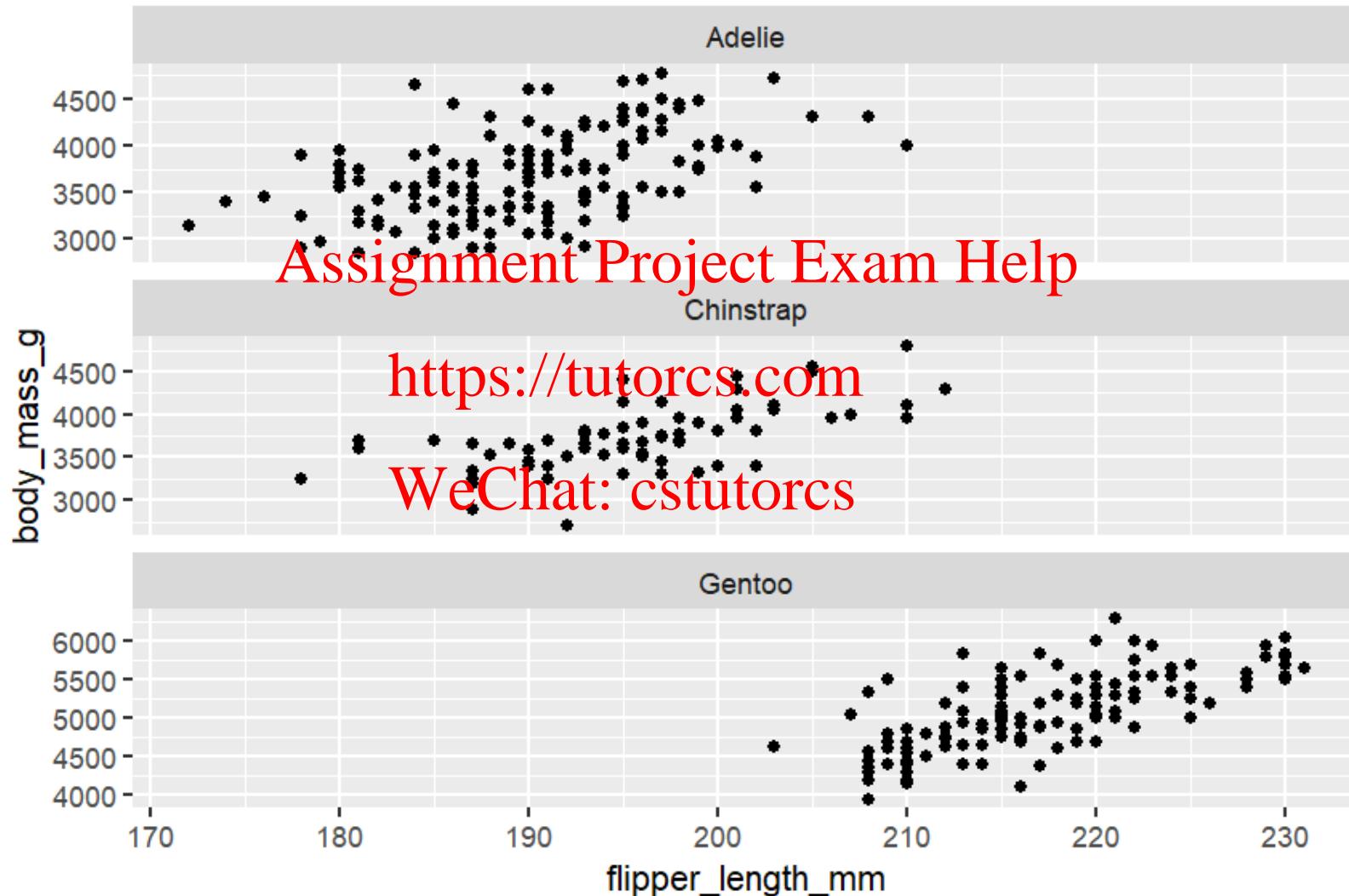
- The number of rows or columns can be changed with the `nrow` or `ncol` arguments

```
ggplot(data = penguins,  
       mapping = aes(x=flipper_length_mm, y=body_mass_g))+  
  geom_point() +  
  facet_wrap(~species, scales = 'free_y', ncol = 1)
```

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

WeChat: cstutorcs

Changing number of columns



Facet grid

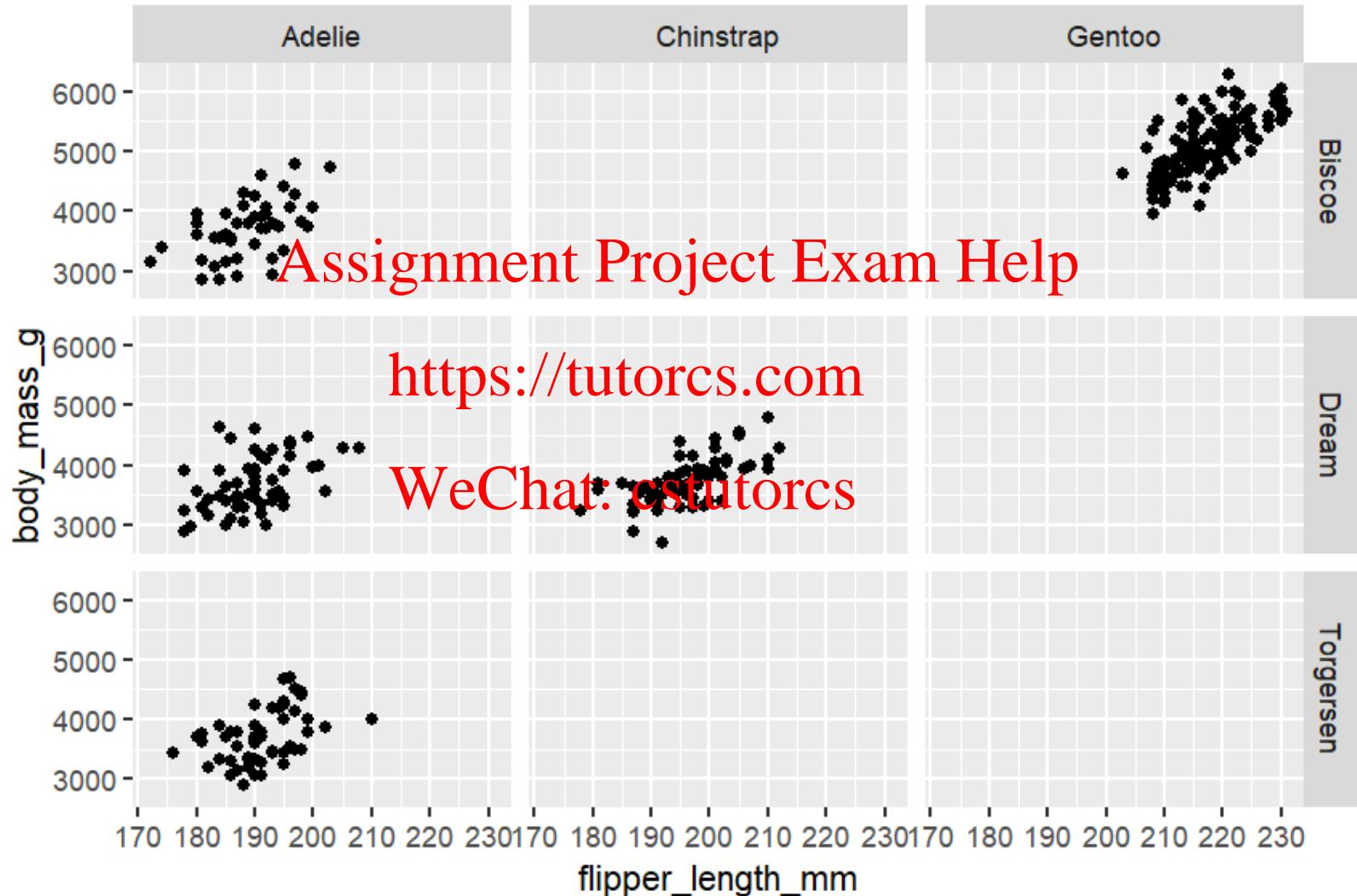
- We can also facet so that the rows correspond to one categorical variable and the columns to another.

```
ggplot(data = penguins,  
       mapping = aes(x=flipper_length_mm, y=body_mass_g))+  
  geom_point() +  
  facet_grid(island~species)
```

<https://tutorcs.com>

WeChat: cstutorcs

Facet grid



Your Turn

- Plot a scatterplot with
 - Bill length on the x axis
 - Bill depth on the y axis
 - Facet by year on the rows
 - Facet by island in the columns
 - Colour by the species

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Solution

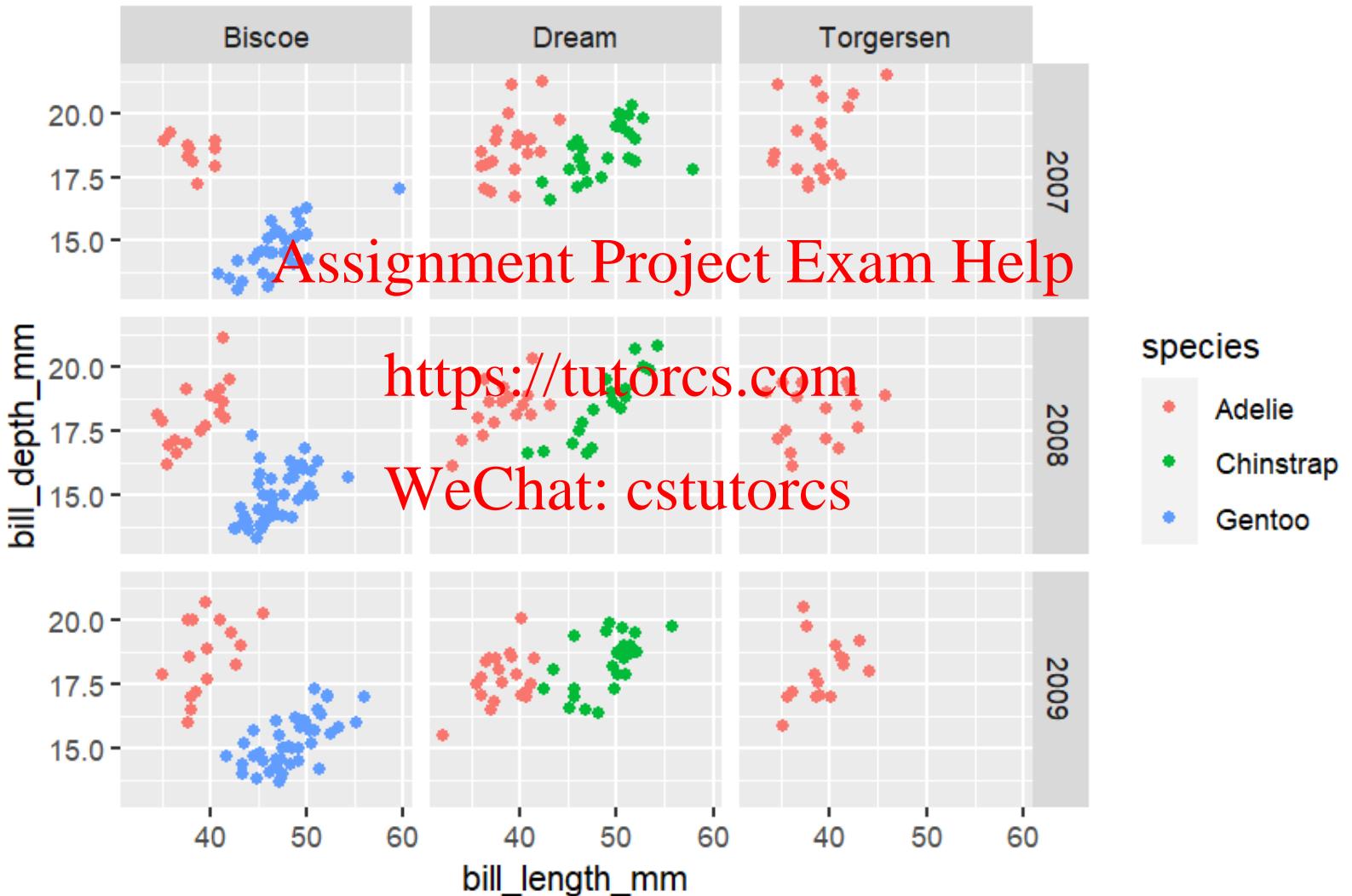
```
ggplot(data = penguins,  
       mapping = aes(x=bill_length_mm, y=bill_depth_mm, colour = species))+  
  geom_point() +  
  facet_grid(year~island)
```

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Solution



Assignment Project Exam Help

Higher Dimensions

WeChat: cstutorcs

<https://tutorcs.com>

Pairs plot

- A pairs plot gives an array of plots
 - On the diagonal there are kernel densities or barplots
 - On the lower diagonal are scatterplots or faceted histograms
 - On the upper diagonal are correlations or boxplots.
- This can be implemented using the `ggpairs` function in the `GGally` package.

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Palmer data

```
library(GGally)  
ggpairs(penguins)
```

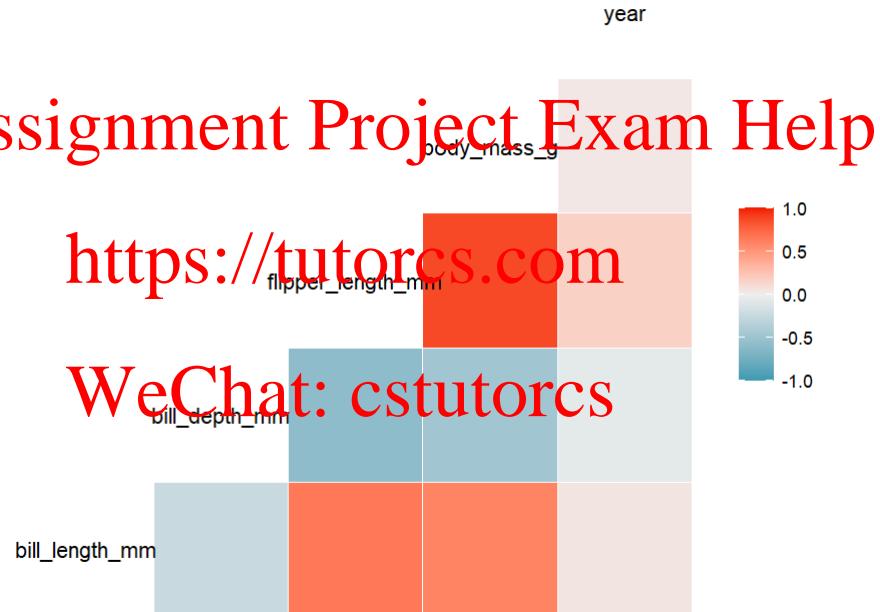
Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Correlation plot

```
ggcorr(penguins)
```



Parallel Coordinates

- A parallel coordinates plots the variables of all values along the y axis.
- The variables themselves appear along the x axis.
- Values corresponding to the same observation are joined up by lines.
- They can often look messy but sometimes provide insight.

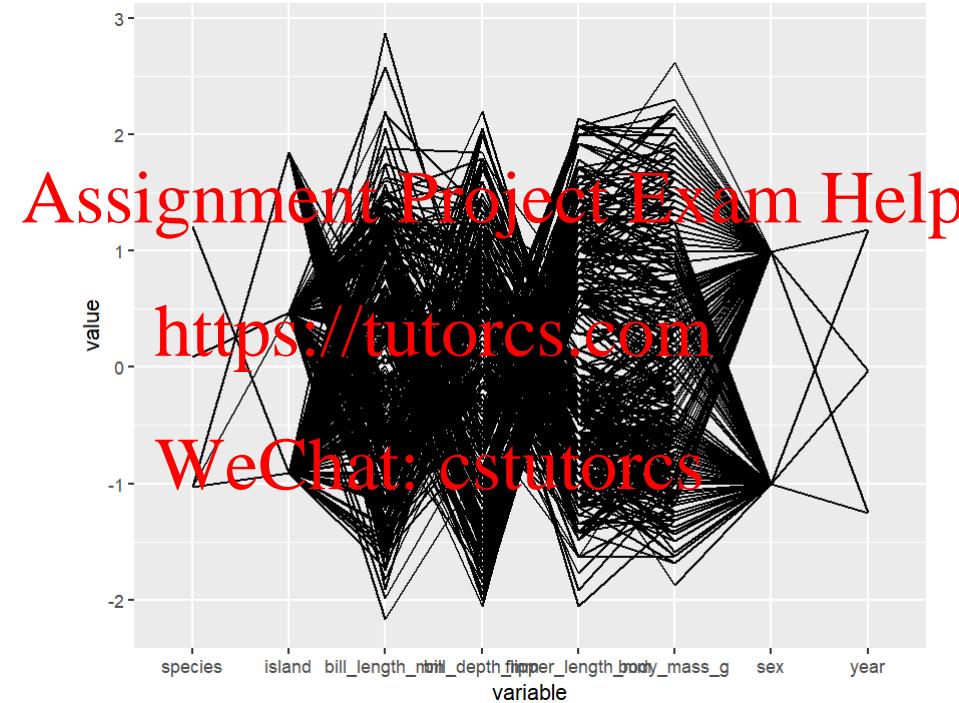
Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Parallel Coordinates

```
ggparcoord(penguins)
```





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 5

