

Assignment Project Exam Help

Lecture 3 : GGR376

Spatial Data Science II

<https://powcoder.com>

Dr. Adams

Add WeChat powcoder

# Assignment Project Exam Help

- ▶ Skewness & Kurtosis
- ▶ Visualization
- ▶ ggplot2

<https://powcoder.com>

Add WeChat powcoder

# Assignment Project Exam Help

<https://powcoder.com>

Skewness and Kurtosis

## Add WeChat powcoder

# Assignment Project Exam Help

- ▶ Rightward / Positive Skew
  - ▶ Long tail of high value numbers
  - ▶  $\text{mean} > \text{median} > \text{mode}$

- ▶ Leftward / Negative Skew

- ▶ Long tail of low value numbers
- ▶  $\text{mean} < \text{median} < \text{mode}$

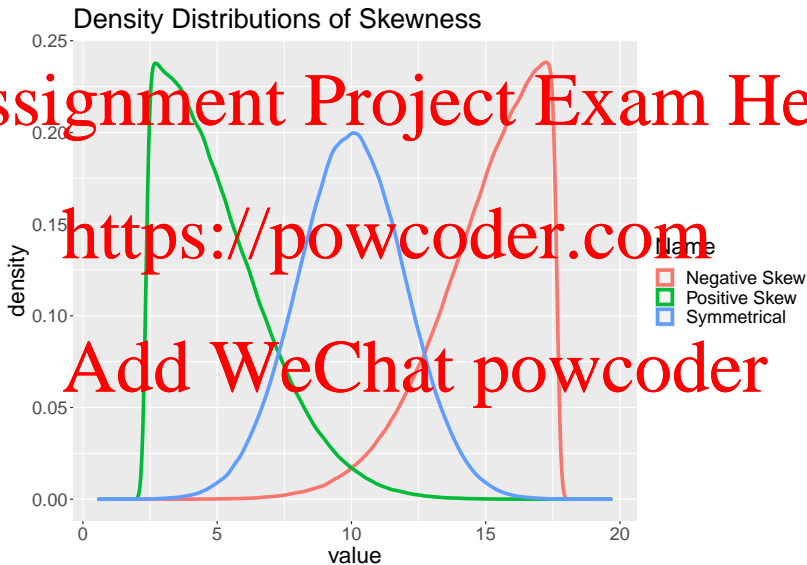
- ▶ Symmetric

- ▶ When there is no skewness in the data.

<https://powcoder.com>

Add WeChat powcoder

# Skewness Visual



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Assignment Project Exam Help

Pearson's moment coefficient of skewness

$$Sk = \frac{\sum(x_i - \bar{x})^3 / N}{Var(x)^{1.5}}$$

<https://powcoder.com>

Many skewness formulas exist, generally though:

- ▶  $Sk = 0$ , Symmetrical
- ▶  $Sk < 0$ , Negative Skewness
- ▶  $Sk > 0$ , Positive Skewness

Add WeChat powcoder

# Assignment Project Exam Help

```
# Need to load a package with a skewness function
```

```
library(moments)
```

```
moments::skewness()
```

<https://powcoder.com>

Add WeChat powcoder

# Assignment Project Exam Help

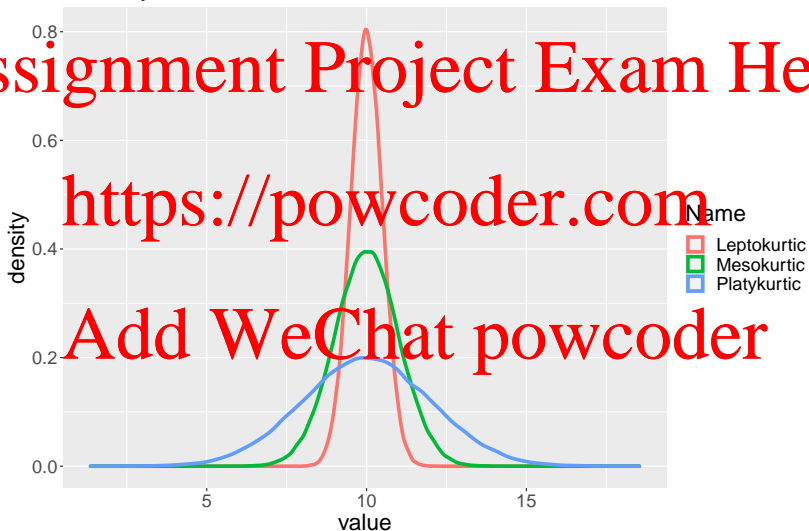
- ▶ Sharpness of the peak of a frequency-distribution curve
- ▶ Leptokurtic
  - ▶ High central peak
- ▶ Mesokurtic
  - ▶ Medium central peak
  - ▶ Standard normal curve
- ▶ Platykurtic
  - ▶ Low central peak

<https://powcoder.com>

Add WeChat powcoder



Density Distributions of Kurtosis



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Assignment Project Exam Help

Pearson's measure of kurtosis

$$Ku = \frac{\sum (x_i - \bar{x})^4 / N}{Var(x)^2}$$

<https://powcoder.com>

Units are not meaningful (as with skewness):

- ▶  $Ku \sim 0$ , Mesokurtic
- ▶  $Ku \geq 0$  and up to infinity, Leptokurtic
- ▶  $Ku < 0$  to -2.75, Platykurtic

Add WeChat powcoder

# Assignment Project Exam Help

```
# Need to load a package with a kurtosis function
```

```
library(moments)
```

```
moments::kurtosis()
```

<https://powcoder.com>

Add WeChat powcoder

# Assignment Project Exam Help

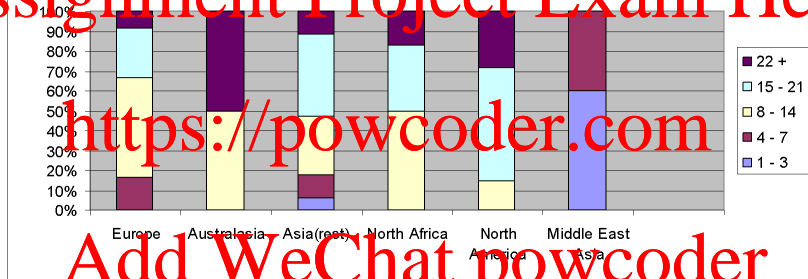
<https://powcoder.com>

Add WeChat powcoder

No more ugly plots.

Assignment Project Exam Help

<https://powcoder.com>



Add WeChat powcoder

[markedbyteachers.com/](https://markedbyteachers.com/)

# Assignment Project Exam Help

The simple graph has brought more information to the data analyst's mind than any other device.

- John Tukey

<https://powcoder.com>

## Add WeChat powcoder

# Assignment Project Exam Help

What makes a memorable (effective) visualization?

(M. A. Borkin et al. 2015)

1. Titles and supporting text need to convey the message of a visualization.
2. Pictograms may be included to improve recognition, unlikely to hinder understanding.
3. Redundancy helps effectively communicate the message.

Add WeChat powcoder

“A memorable visualization is often also an effective one.”

# Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



# Assignment Project Exam Help

**The relationship between alcohol use and long-term cognitive decline in middle and late life: a longitudinal analysis using UK Biobank**

Giovanni Piumatti, Simon C Moore, Damon M Berridge, Chinmoy Sarkar, John Gallacher

*Journal of Public Health*, <https://doi.org/10.1093/pubmed/fox186>

<https://powcoder.com>

Add WeChat powcoder

# Assignment Project Exam Help

Explain the causal (not correlation) between cognitive decline and alcohol intake in middle and older aged adults in the UK.

<https://powcoder.com>  
Add WeChat powcoder

# Assignment Project Exam Help

- ▶ UK Department of Health, drinkers should not consume more than 16 g/day to minimize the risk of alcohol to health.

Add WeChat powcoder

- ▶ Data from 13,342 men and women, aged between 40 and 73 years.
- ▶ How often and how much alcohol was consumed.
- ▶ Regression analysis testing the functional relationship and impact of alcohol on cognitive performance.
- ▶ Performance was measured in response of card matching (do a pair of cards match?).
- ▶ Covariates included body mass index, physical activity, tobacco use, socioeconomic status, education and baseline cognitive function.
  - ▶ Additional variables that may be predictive of the outcome
    - ▶ Direct interest or it may be a confounding or interacting variable

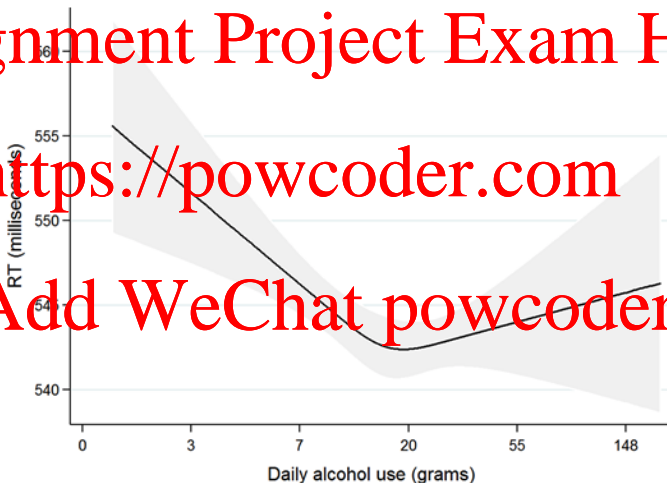
# Assignment Project Exam Help

“UK department of Health guidelines are that drinkers should not consume more than 16 g/day to minimize the risk of alcohol to health. Our findings suggest that to preserve cognitive performance **10 g/day** is a more appropriate upper limit.”

<https://powcoder.com>  
Add WeChat powcoder

Results - To preserve cognitive performance 10 g/day limit

Fig. 1



# Assignment Project Exam Help

“What is a graphic? How can we succinctly describe a graphic?  
And how can we create the graphics that we have described?”  
(Wickham, 2010)

<https://powcoder.com>

One answer to these questions was to develop a grammar of graphics. - Grammar referring to a set of rules or constructs to govern the process.

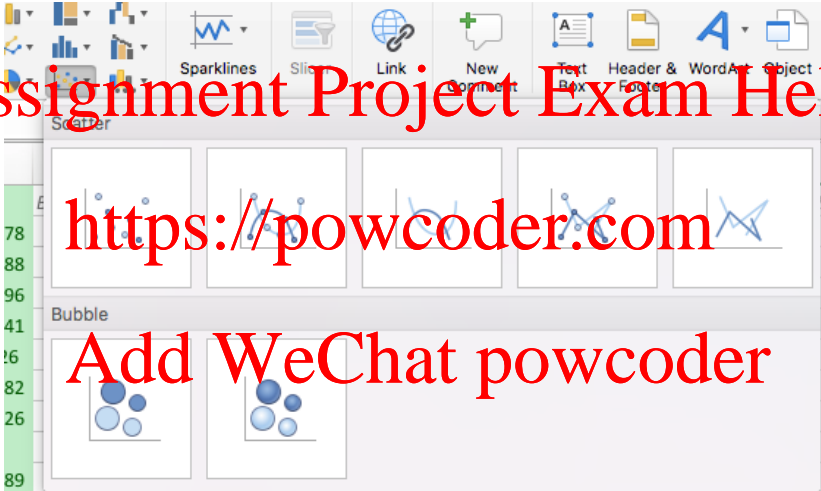
Add WeChat powcoder

What is different about the grammar of graphics?

Assignment Project Exam Help

<https://powcoder.com>

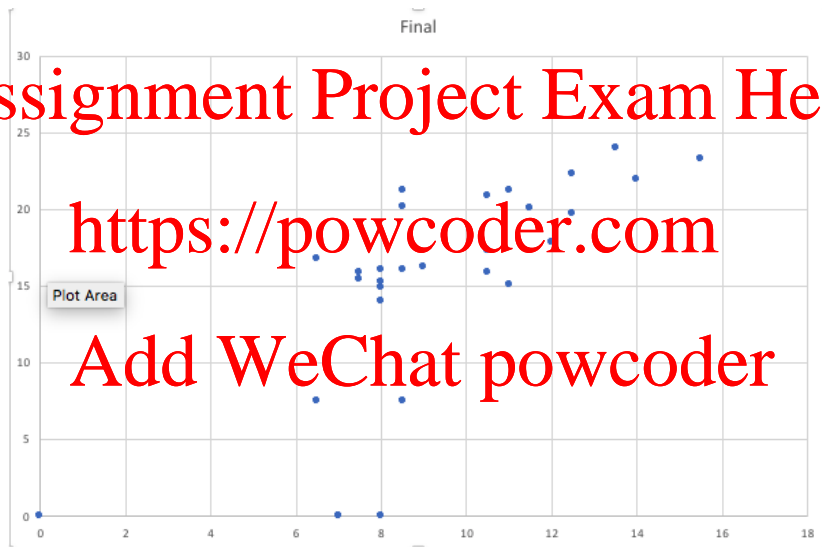
Add WeChat powcoder



Traditionally, you will select the plot by name.



Produce a Plot



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

And try your best to modify the plot

Select Data Source

**Range Details**

Chart data range:

**Legend entries (Series):**

**Final**

Name:

X values:

Y values:

Horizontal (Category) axis labels:

**Hidden and Empty Cells**

Show empty cells as:

☐ Show data in hidden rows and columns

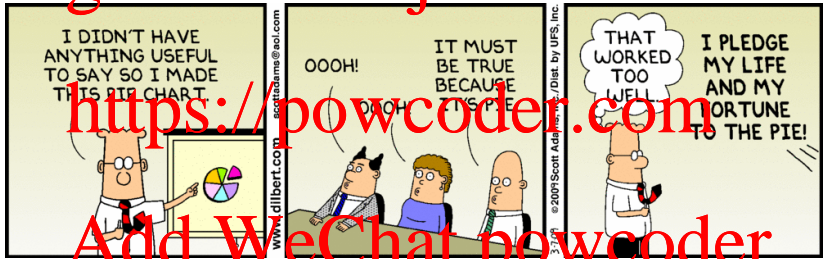
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

We can do better

# Assignment Project Exam Help



# Assignment Project Exam Help

- ▶ A dataset and the mapping of those data to aesthetics
- ▶ At least one layer, including:

<https://powcoder.com>

- ▶ geometric object
- ▶ statistical transformation
- ▶ A scale for each aesthetic mapping
- ▶ A coordinate system for the plot
- ▶ Faceting

Add WeChat powcoder

How are plots constructed?

# Assignment Project Exam Help

Using the grammar of graphics we construct our plots one component at a time.

<https://powcoder.com>

Add WeChat powcoder

# Assignment Project Exam Help

Using a grammar of graphics, we can adjust each piece in isolation.

<https://powcoder.com>

Add WeChat powcoder

## Gapminder example

gapminder: Data from Gapminder

An excerpt of the data available at Gapminder.org. For each of 142 countries, the package provides values for life expectancy, GDP per capita, and population, every five years, from 1952 to 2007.

```
library(gapminder)
colnames(gapminder::gapminder)[1:3]
```

```
## [1] "country" "continent" "year"
```

```
colnames(gapminder::gapminder)[4:6]
```

```
## [1] "lifeExp" "pop" "gdpPercap"
```

Subset the data to 2002

# Assignment Project Exam Help

```
gap <- garbinder::garbinder %>%  
  dplyr::filter(year == 1997)
```

Add WeChat powcoder



## Review the subset

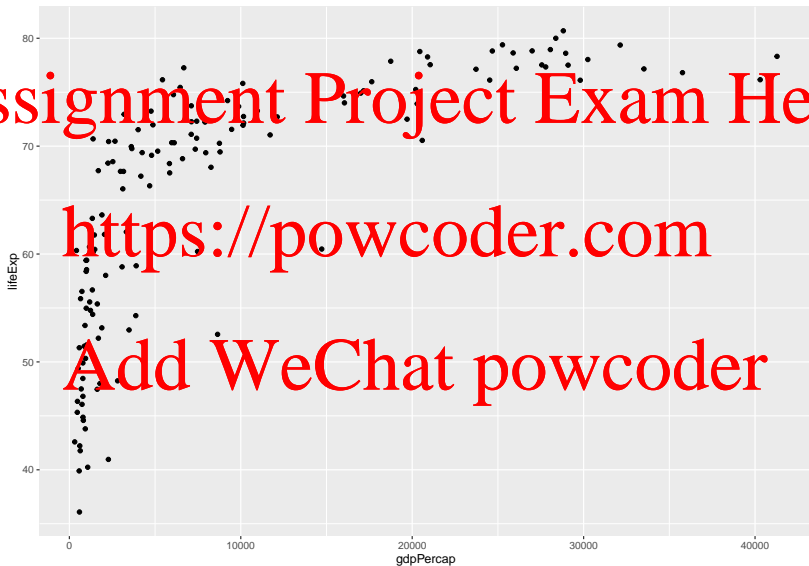
continent	year	lifeExp	pop
Africa :52	Min. :1997	Min. :36.09	Min. :1.456
Americas:25	1st Qu.:1997	1st Qu.:55.63	1st Qu.:3.770
Asia :33	Median :1997	Median :69.39	Median :9.735
Europe :30	Mean :1997	Mean :65.01	Mean :3.884
Oceania :2	3rd Qu.:1997	3rd Qu.:74.17	3rd Qu.:2.431
	Max. :1997	Max. :80.69	Max. :1.230

gdpPercap

Min. :	312.2
1st Qu. :	1366.8
Median :	4781.8
Mean :	9090.2
3rd Qu. :	12022.9
Max. :	41283.2

Add WeChat powcoder

## Examine Life Expectancy and GDP



# Assignment Project Exam Help

```
ggplot(data = <data>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

Add WeChat powcoder

# Assignment Project Exam Help

- ▶ Geometric objects include points, lines, and polygons (similar to our vector data model).

The grammar of graphics requires that you specify the geometric feature that will be used to render the plot.

In ggplot, they are known as `geoms` (geometric objects).

<https://powcoder.com>

Add WeChat powcoder

# Assignment Project Exam Help

- ▶ Every geom is related with a certain statistic.
- ▶ The geom 'histogram' uses the bin statistic
  - ▶ Counts by bins values
- ▶ The same is true for each statistic, it has an associated geom

<https://powcoder.com>  
Add WeChat powcoder

geom\_histogram & stat\_bin

# Assignment Project Exam Help

```
geom_histogram(mapping = NULL, data = NULL, stat = "bin",  
position = "stack", ..., binwidth = NULL, bins = NULL, na.rm =  
FALSE, show.legend = NA, inherit.aes = TRUE)
```

Using <https://powcoder.com> is an alternative approach to creating the layer

```
stat_bin(mapping = NULL, data = NULL, geom = "bar",  
position = "stack", ..., binwidth = NULL, bins = NULL, center =  
NULL, boundary = NULL, breaks = NULL, closed = c("right",  
"left"), pad = FALSE, na.rm = FALSE, show.legend = NA,  
inherit.aes = TRUE)
```

Add WeChat powcoder

mapping = **aes()**

**aes()**

Assignment Project Exam Help

'Aesthetics: things that we can perceive on the graphic.' (Wickham 2010)

- ▶ x-location (x)
- ▶ y-location (y)
- ▶ alpha
- ▶ colour
- ▶ fill
- ▶ group
- ▶ linetype
- ▶ size
- ▶ weight

<https://powcoder.com>

Add WeChat powcoder

**mapping** = aes()

# Assignment Project Exam Help

mapping =

Defining how we associate each element (variable) to an aesthetic (visual element of the plot).

**Add WeChat powcoder**



## One Variable geoms

### Continuous Variables

- ▶ Kernel Density (smooth histogram)

- ▶ `geom_density()`

- ▶ Useful if data come from a smooth distribution

- ▶ Dot plot

- ▶ `geom_dotplot()`

- ▶ Histogram

- ▶ `geom_histogram()`

### Discrete Variable

- ▶ Bar

- ▶ `geom_bar()`

Assignment Project Exam Help

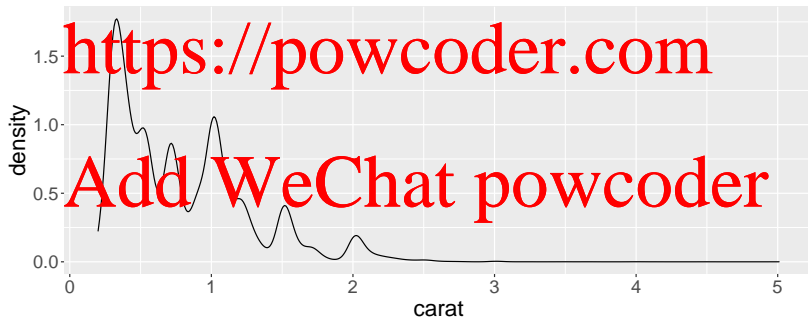
<https://powcoder.com>

Add WeChat powcoder

geom\_density()

```
ggplot(diamonds, aes(carat)) +  
  geom_density() +  
  theme(text = element_text(size=20))
```

Assignment Project Exam Help



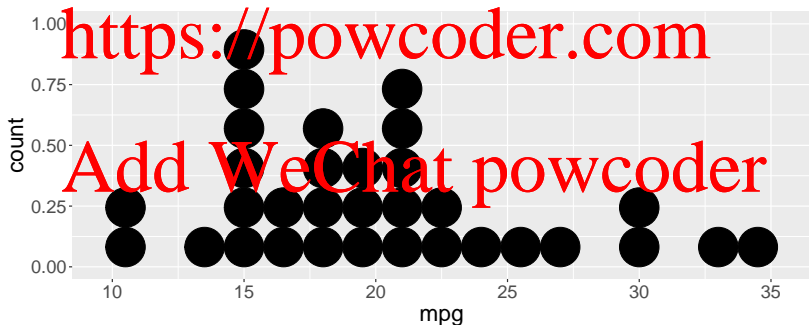
<https://powcoder.com>

Add WeChat powcoder

stat\_density()

geom\_dotplot()

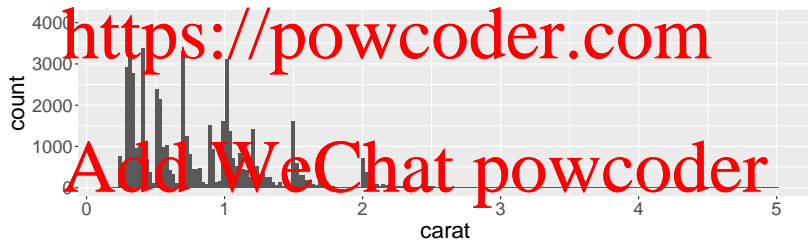
```
# Use fixed-width bins  
ggplot(mtcars, aes(x = mpg)) +  
  geom_dotplot(method="fixedwidth", binwidth = 1.5) +  
  theme(text = element_text(size=20))
```



stat\_identity() - The identity statistic leaves the data unchanged.

geom\_histogram()

```
ggplot(diamonds, aes(carat)) +  
  geom_histogram(bins = 200)  
  theme(text = element_text(size=20))
```

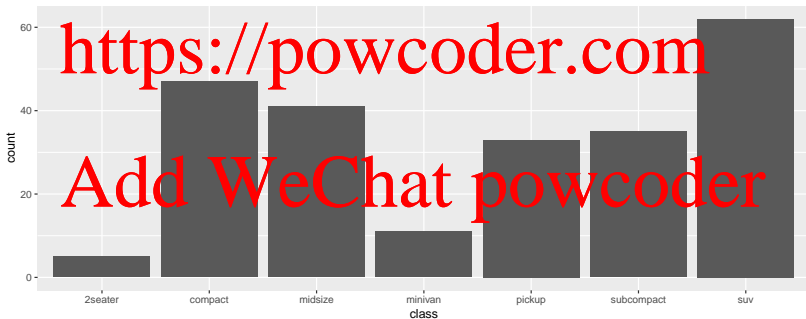


stat\_bin(..., binwidth = NULL, bins = NULL, ...) - Continuous data, count values by bins

geom\_bar()

```
# Car counts per class
```

```
ggplot(mpg, aes(class)) +  
  geom_bar()
```



stat\_count() - Count observations by grouping (discrete)

- ▶ Points

- ▶ `geom_point()`

- ▶ Points with a little noise

- ▶ `geom_jitter()`

- ▶ Quantile lines

- ▶ `geom_quantile()`

- ▶ Performs quantile regression and draws the fitted quantiles with lines

- ▶ Smoothed conditional means

- ▶ `geom_smooth()`

- ▶ Include standard errors

# Assignment Project Exam Help

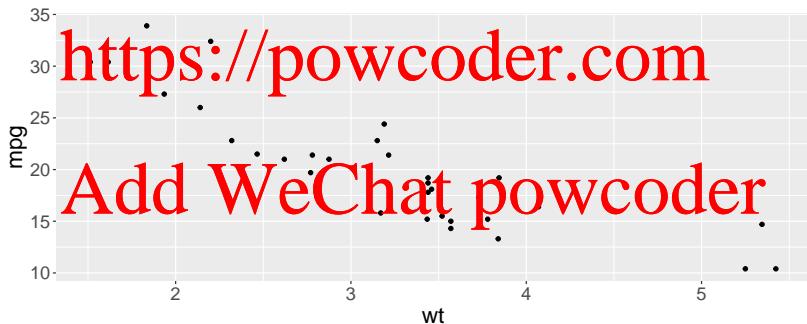
<https://powcoder.com>

Add WeChat powcoder

geom\_point()

```
ggplot(mtcars, aes(wt, mpg))+  
  geom_point()+  
  theme(text = element_text(size=20))
```

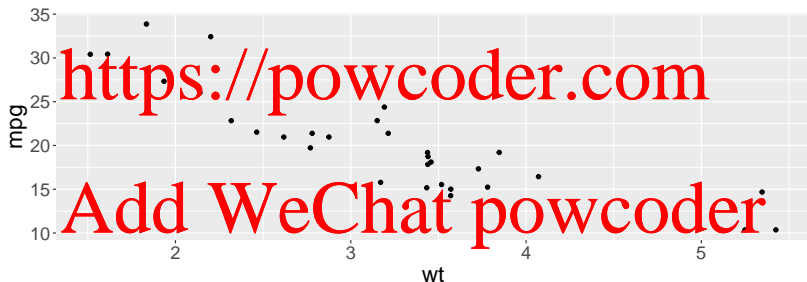
Assignment Project Exam Help



stat\_identity()

geom\_jitter()

```
ggplot(mtcars, aes(wt, mpg))+  
  geom_jitter()+  
  theme(text = element_text(size=20))
```

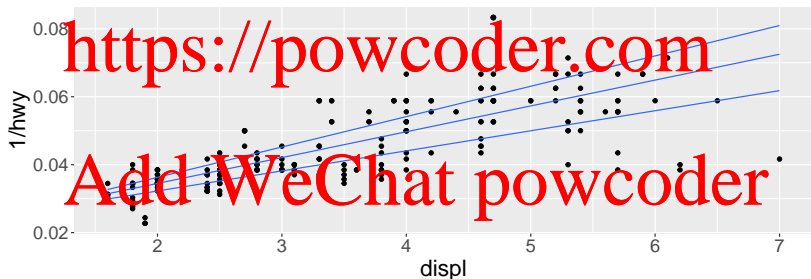


- Used for handling overplotting caused by discreteness in smaller datasets  
stat\_identity()



geom\_quantile()

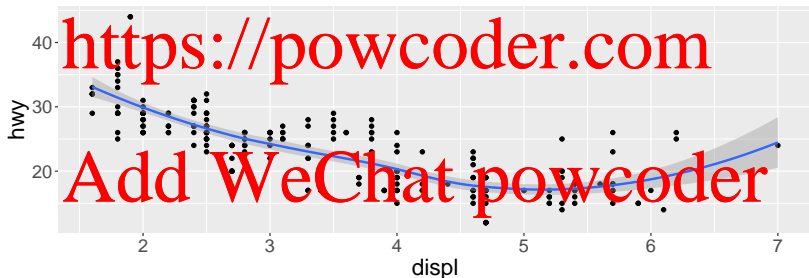
```
ggplot(mpg, aes(displ, 1 / hwy)) + geom_point() +  
  geom_quantile() +  
  theme(text = element_text(size=20))
```



stat\_quantile(..., quantiles = c(0.25, 0.5, 0.75),... )  
- Quantile regression

geom\_smooth()

```
ggplot(mpg, aes(displ, hwy)) + geom_point() +  
  geom_smooth() +  
  theme(text = element_text(size=20))
```



stat\_smooth(..., level = 0.95, ...)  
- Conditional mean

# Assignment Project Exam Help

ggplot requires that we specify either a `geom_*` or a `stat_*` when we construct our plots

<https://powcoder.com>

Add WeChat powcoder

# Assignment Project Exam Help

Scales are used to determine how each aesthetic will appear. -  
Where in graphical space should the points occur? - What is the  
necessary range in colour? - What is the range in size of points?

As you progress you may begin customizing scales, but not at  
this time.

Add WeChat powcoder

# Assignment Project Exam Help

Known as `coord` in `ggplot`, maps the position of objects in graphical space.

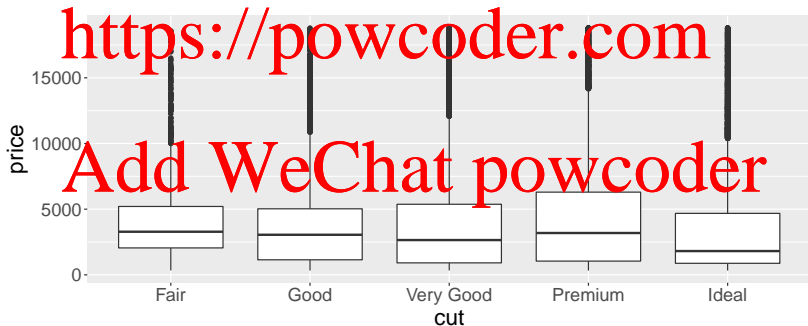
They affect how geoms look in the plot. Changing the `coord` from `coord_cartesian` to `coord_polar` for a `geom_bar` changes the geometry to a circle.

`coord_cartesian()`

Cartesian coordinate system specifies each point uniquely in a plane by a pair of numerical coordinates  $(x, y)$ , which are the signed distances to the point from two fixed perpendicular directed lines  $(0,0)$ , measured in the same unit of length.

geom\_boxplot()

```
ggplot(diamonds, aes(cut, price)) +  
  geom_boxplot() +  
  theme(text = element_text(size=20))
```

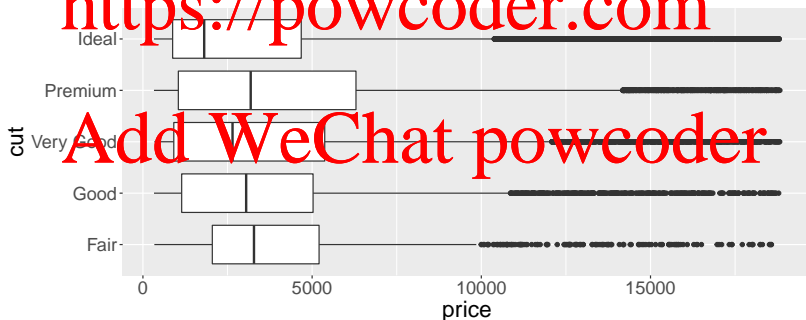


geom\_boxplot()+coord\_flip()

```
ggplot(diamonds, aes(cut, price))+  
  geom_boxplot()+  
  coord_flip()+  
  theme(text = element_text(size=20))
```

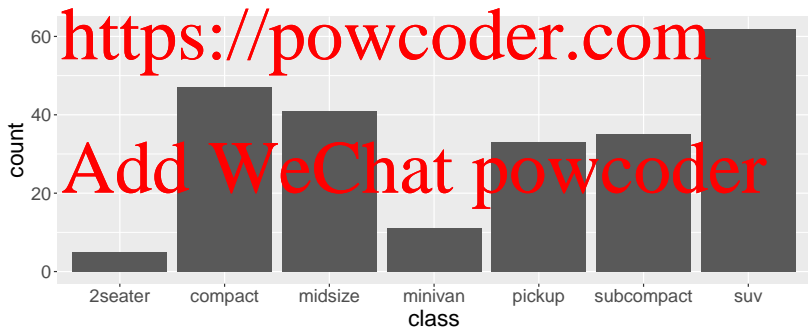
Assignment Project Exam Help

<https://powcoder.com>



geom\_bar

```
ggplot(mpg, aes(class)) +  
  geom_bar() +  
  theme(text = element_text(size=20))
```





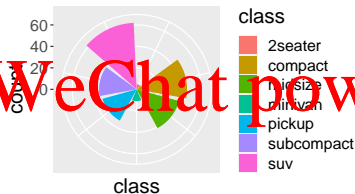
```
geom_bar + coord_polar()
```

# Assignment Project Exam Help

```
ggplot(mpg) +  
  geom_bar(mapping = aes(x = class, fill = class)) +  
  theme(axis.text.x = element_blank(), text = element_text(  
    coord_polar()
```

<https://powcoder.com>

Add WeChat powcoder



# Assignment Project Exam Help

Three variables is the maximum recommended number of variables that should be visualized with a single plot.

<https://powcoder.com>

- ▶ Two variables visualized by position.
- ▶ One variable by another aesthetic

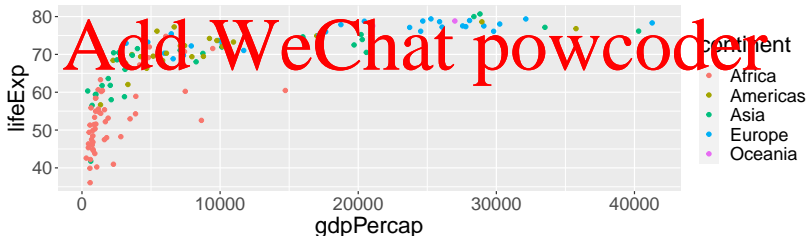
Add WeChat powcoder

## More than three leads us to using facets

- ▶ Creating subplots
- ▶ Facet based on one variable, `facet_wrap()`
- ▶ Facet based on two variables, `facet_grid()`

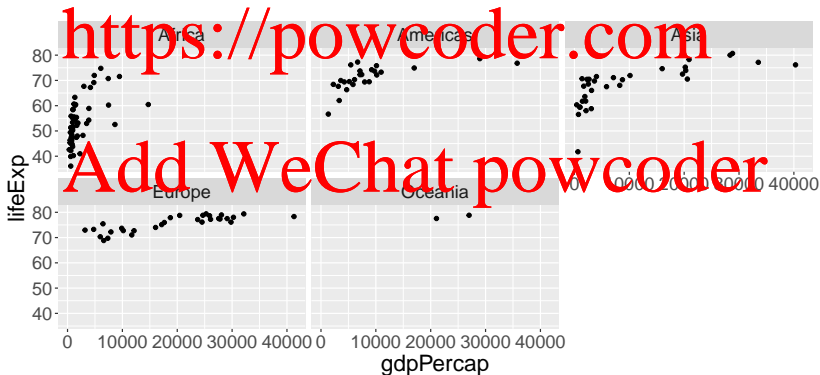
Assignment Project Exam Help

```
ggplot(data = gap)+  
  geom_point(mapping = aes(gdpPerCap, lifeExp,  
                           colour = continent))+ theme(text = size(12))
```



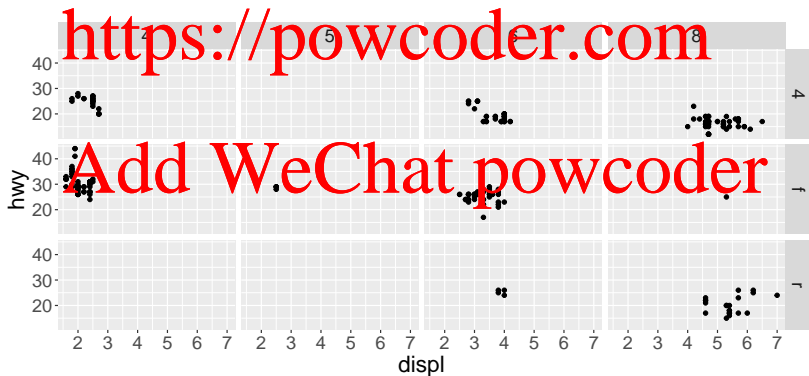
```
facet_wrap()
```

```
ggplot(data = gap)+  
  geom_point(mapping = aes(gdpPercap, lifeExp))+  
  facet_wrap(~continent, nrow = 2)+  
  theme(text = element_text(size=20))
```



facet\_grid()

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy)) +  
  facet_grid(drv ~ cyl) +  
  theme(text = element_text(size=20))
```



# Assignment Project Exam Help

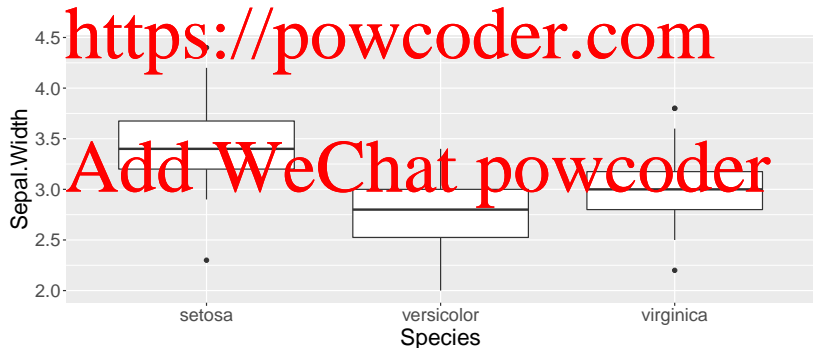
- ▶ Too many variables, which can be overcome through facets
- ▶ Overplotting (discrete outcomes), we can jitter the values
- ▶ Alphabetical ordering, we can define our order using factors or by mean/median values
- ▶ Polar coordinates, you must be careful when using these as humans struggle during interpretation.

<https://powcoder.com>  
Add WeChat powcoder

## Ordering demo

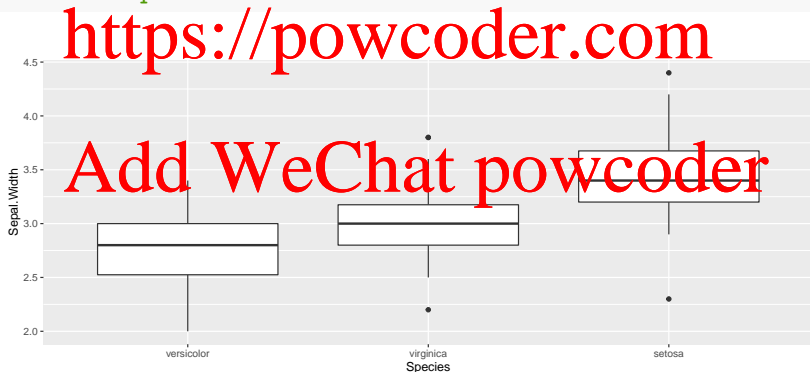
```
# Default order is alphabetical
```

```
geomplot(iris, aes(x = Species, y = Sepal.Width))  
geom_boxplot()  
theme(text = element_text(size=20))
```



## Order by median

```
ggplot(iris, aes(x =  
  reorder(Species, Sepal.Width, FUN = median),  
  y = Sepal.Width))  
  geom_boxplot()+  
  xlab("Species")
```





# Assignment Project Exam Help

Read: Chapters 2 & 3, H4DS  
<https://powcoder.com>

Add WeChat powcoder

## References

Borkin, Michelle A, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, Senior Member, and Aude Oliva. 2015. "Beyond Memorability: Visualization Recognition an," 1–10. papers3://publication/uuid/9AD97C9C-BAC3-414A-8B6E-G951EE2CB7FB.

Groeneveld, RA, and G Meeden. 1984. "Measuring Skewness and Kurtosis." *Journal of the Royal Statistical Society. Series D (the Statistician)* 33 (4): 391–99.

Wickham, Hadley. 2010. "A Layered grammar of graphics." *Journal of Computational and Graphical Statistics* 19 (1): 3–28. doi:10.1198/jcgs.2009.07098.