# Quantitative social science with R

## Visualisation and analysis of data

Edu Gonzalo Almorox

18/10/2017

# Quantitative social science with R

## Visualisation and analysis of data

Edu Gonzalo Almorox

18/10/2017

# Outline

1. Visualise data 
    - quantitative data
    - qualitative data
2. Analyse data
    - tidy
    - augment

# ggplot2

## Quantitative data

# ggplot2

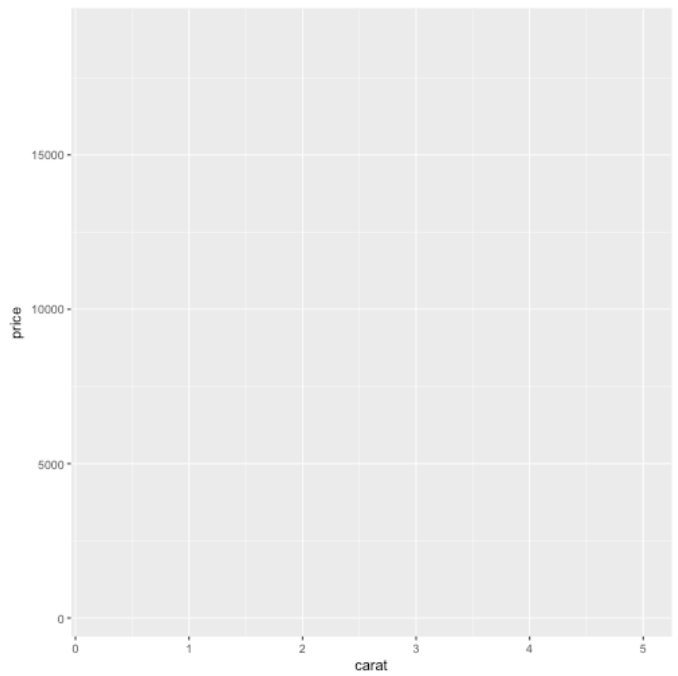
- is one of the most developed packages in R
- It is based on the grammar of graphics

| | |
|---|---|
| Data | The raw data that you want to plot |
| Geometries | The geometric shapes that will represent the data. |
| Aethetics | Aesthetics of the geometric and statistical objects, such as color, size, shape and position. |
| Scales | Maps between the data and the aesthetic dimensions, such as data range to plot width or factor values to colors |

- It is important to have a clear idea of how the data are structured

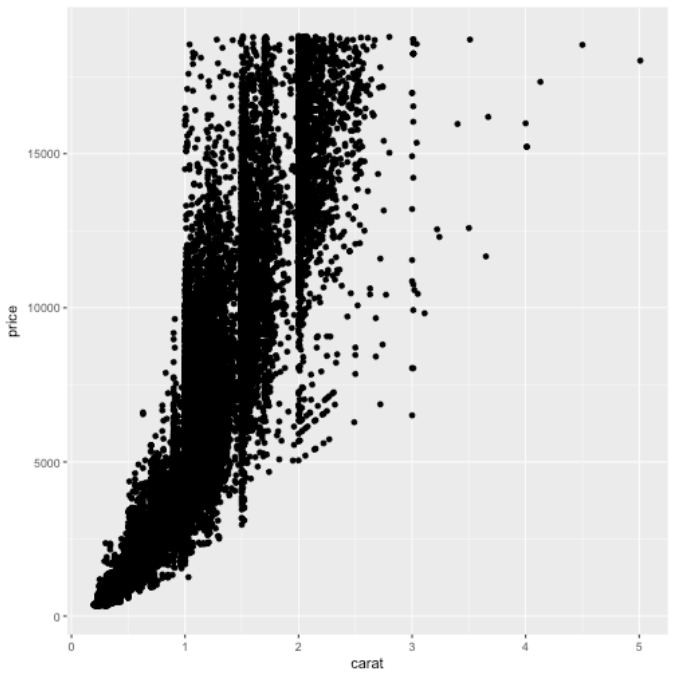
# Hello world with ggplot2

- Compare two variables: e.g. the carat (  ) and price of the diamonds



# Hello world with ggplot2

- Add more layers to define the type of plot

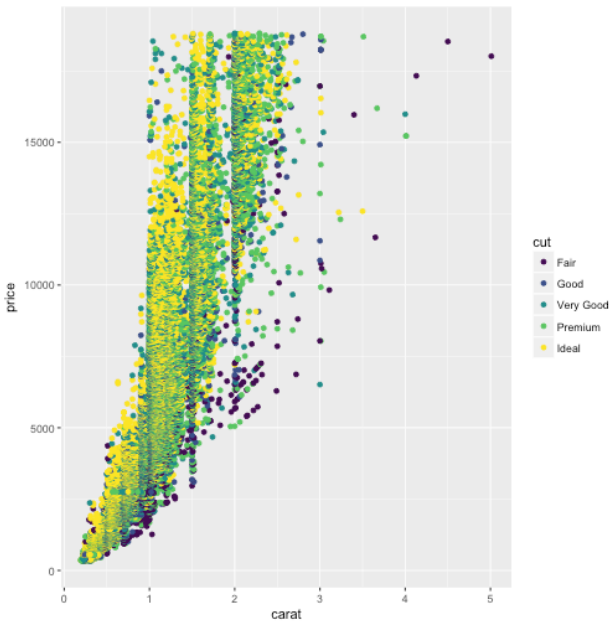


# Hello world with ggplot2

- Can we improve this plot?...

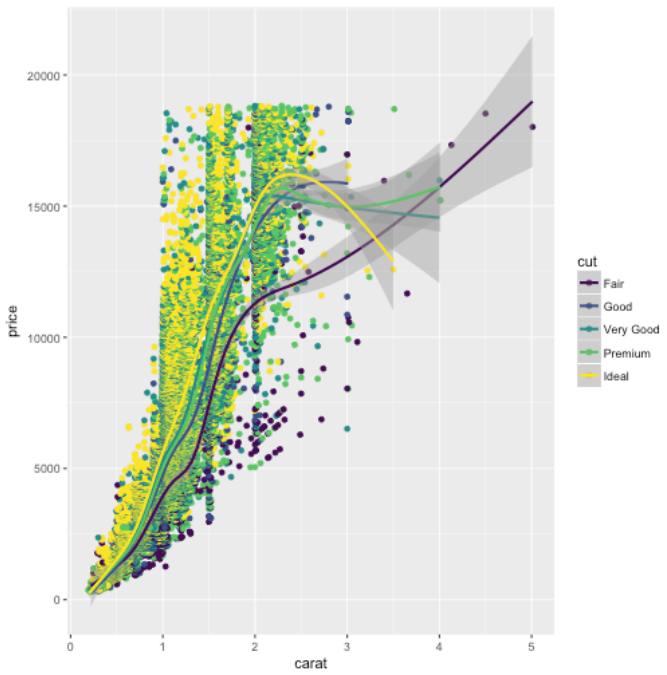
# Hello world with ggplot2

- Can we improve this plot?...



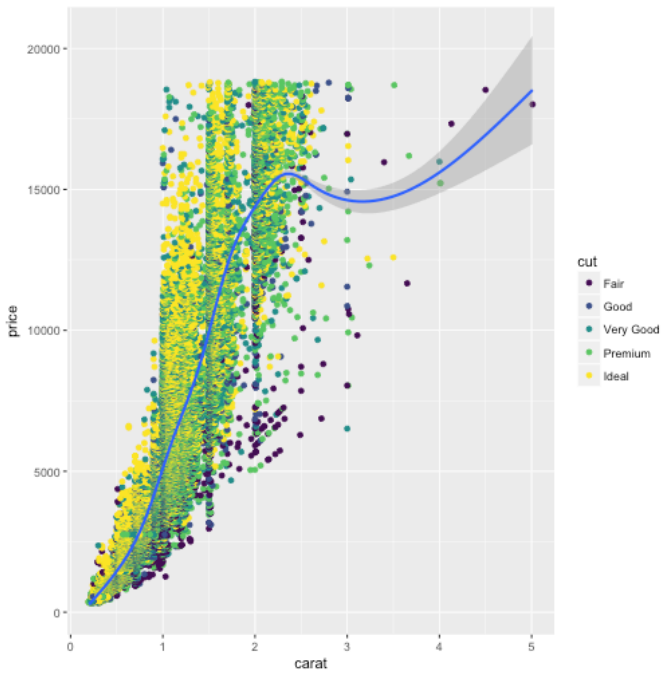
# Hello world with ggplot2

- We can add more layers



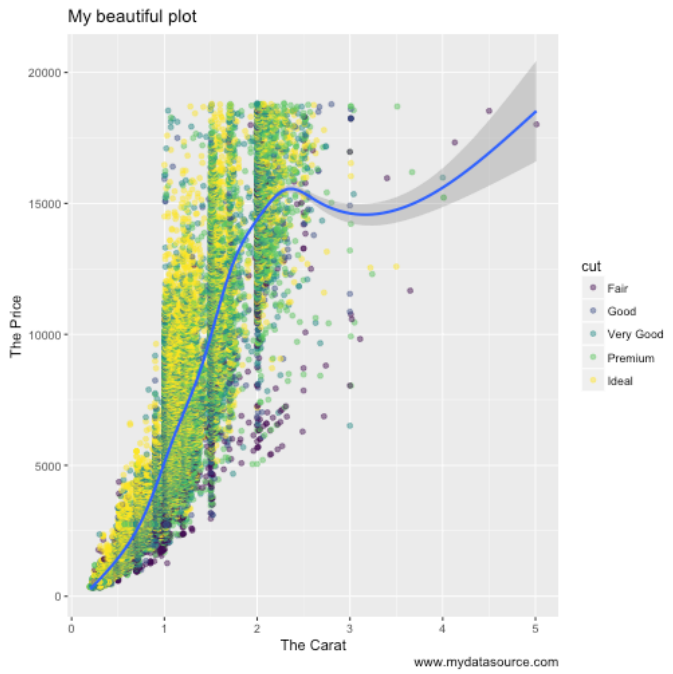
# Hello world with ggplot2

- … and adapt them



# Refinements

- Transparency and labels

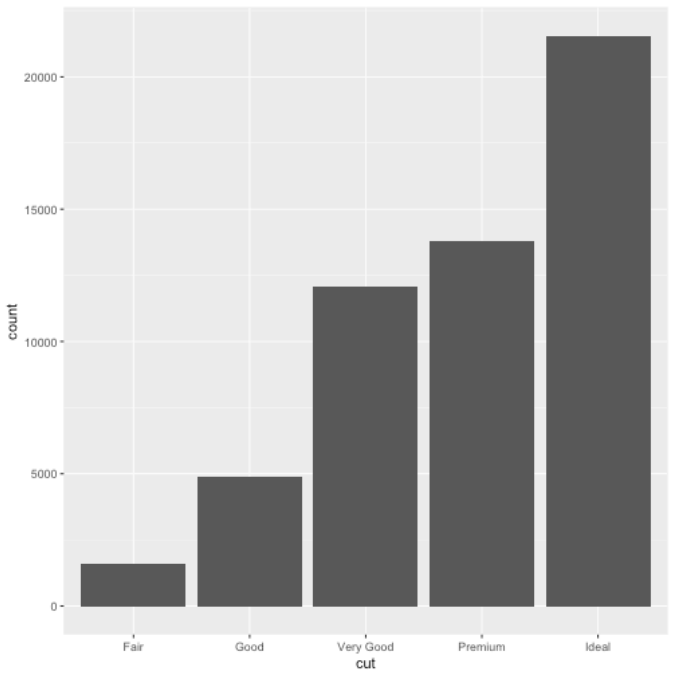


# ggplot2

## Qualitative data

# Barplots

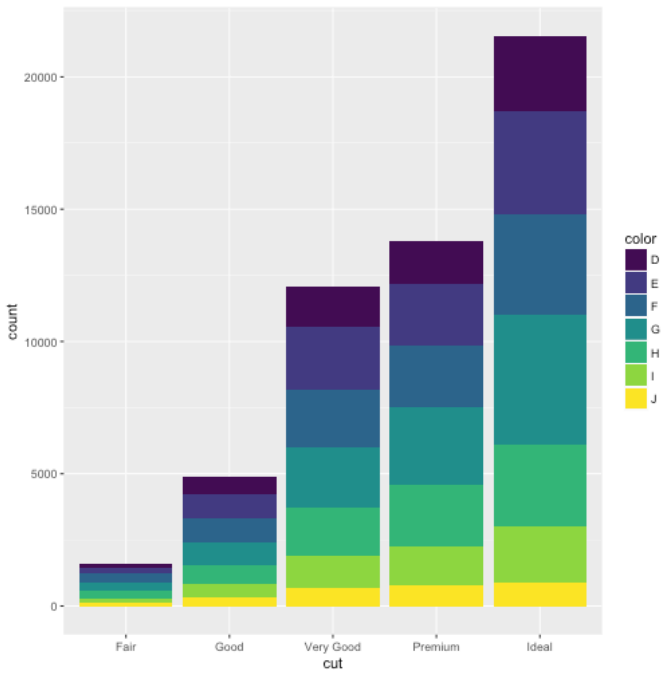
- How many observations of each cut?




- : try to run . What happens?

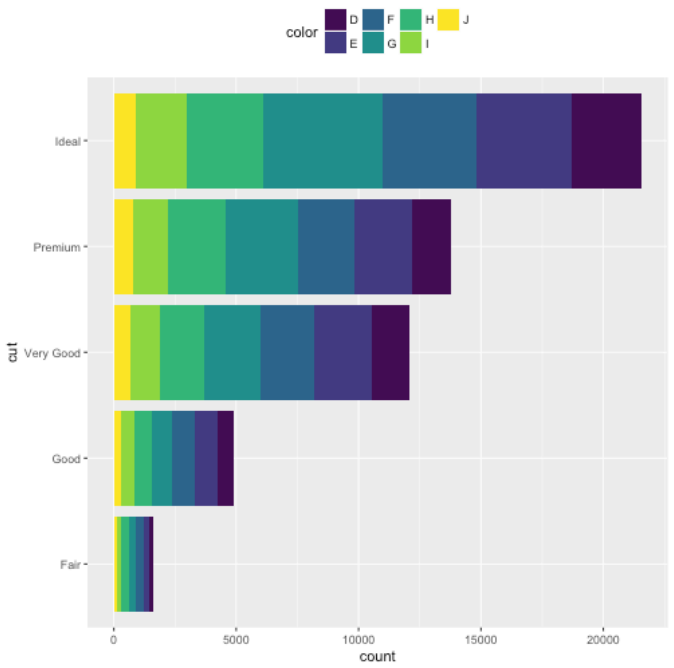
# Barplots

- How many observations of each cut and each colour?



# Barplots

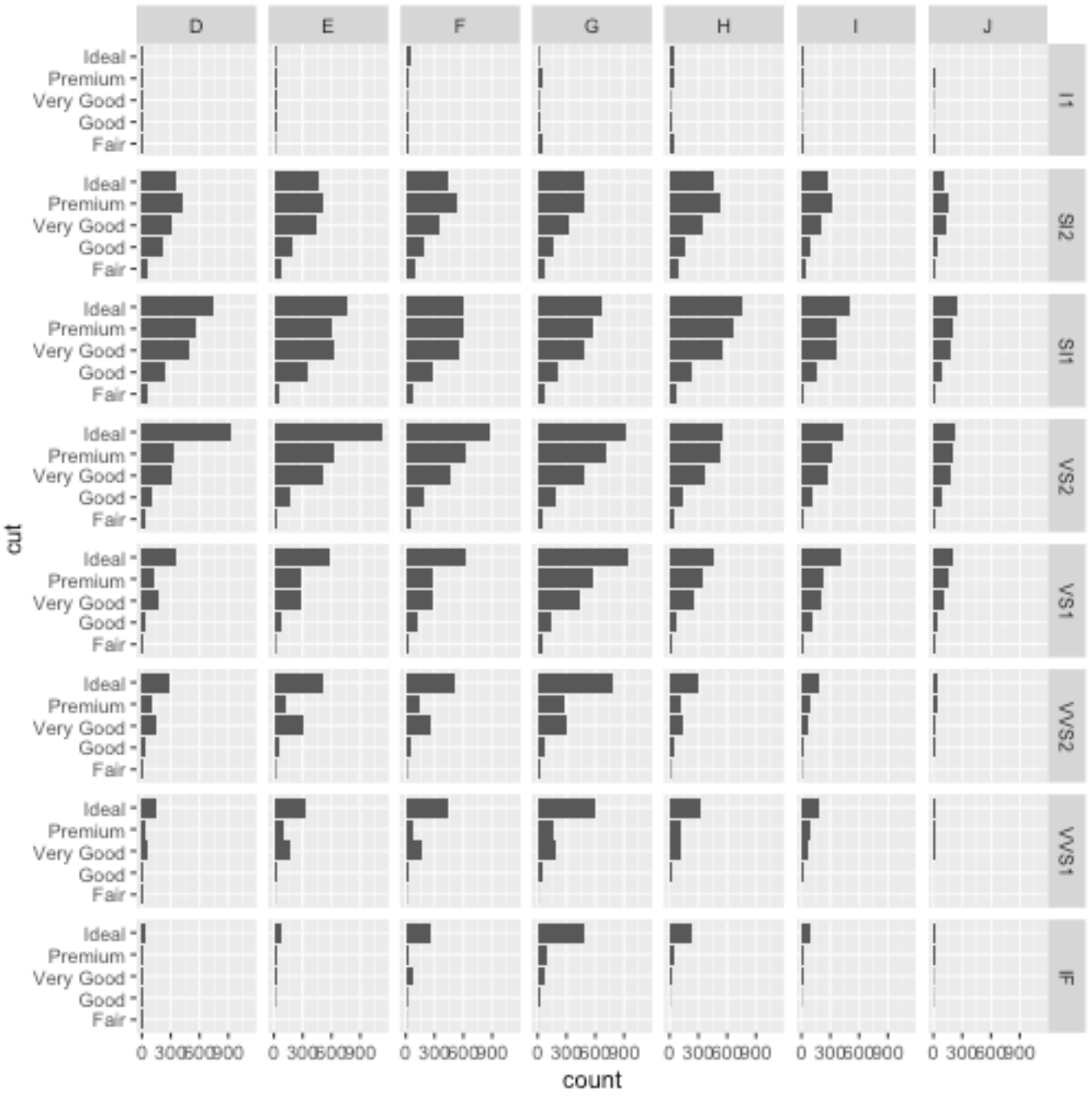
- Can we represent this information in a different way?



# Barplots

- Can we consider more relationships with more variables - e.g. the clarity of the diamonds?

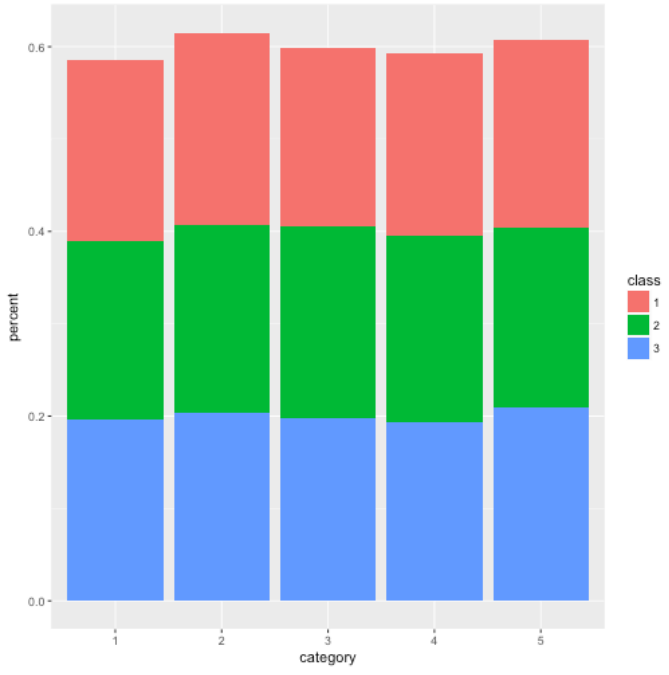
# Barplots


D
E
F
G
H
I
J
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF
cut
Ideal
Premium
Very Good
Good
Fair
0 300600900
count

# Case study

How can we represent the figure below using
?

- format of the data
- current structure of the data
- desired structure of the data
- format of the variables



# Case study

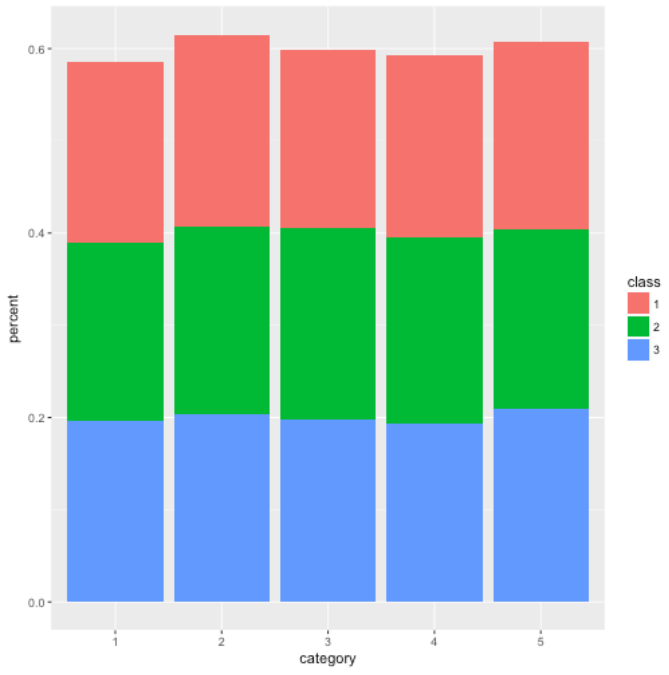
- Format and structure of the data

# Case study

- Prepare the data to get a better structure for visualisation

# Case study

- Visualise data



# Analyse data

## broom

# broom

- broom creates tidy data frames derived from the results of your models
- Focus on the output of the model
- Great compatibility with dplyr functions

| | |
|---|---|
| tidy() | component-level statistics |
| augment() | observation-level statistics |
| glance() | model-level statistics |

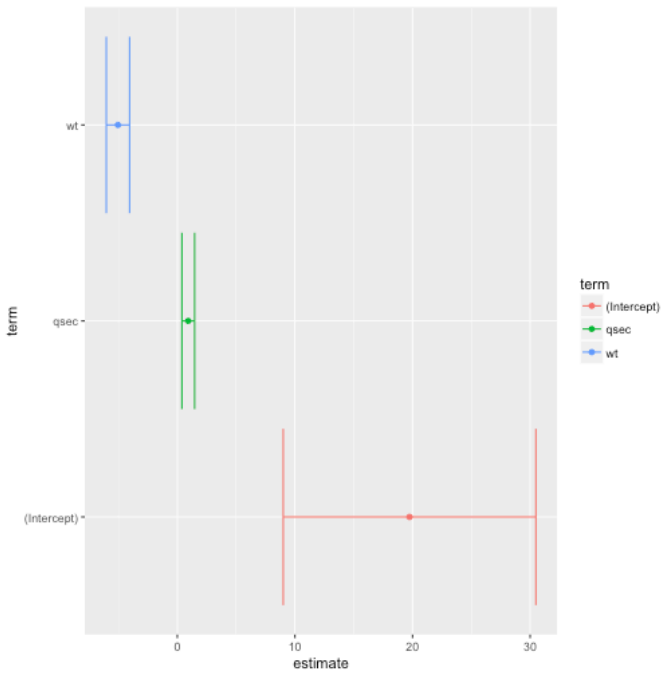
# broom

# broom

- Get tidy results of your model

# broom

- Visualise results



# broom

- Get information referred to the observations

# Thanks!

@EdudinGonzalo