ODDS AND ENDS

Overview

Very, brief history and discussion of data visualizations

Additional ggplot features and visual hypothesis test

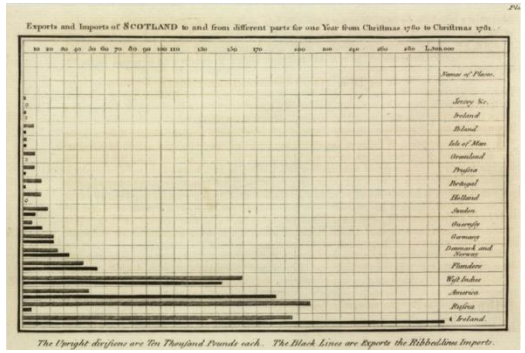
Writing functions

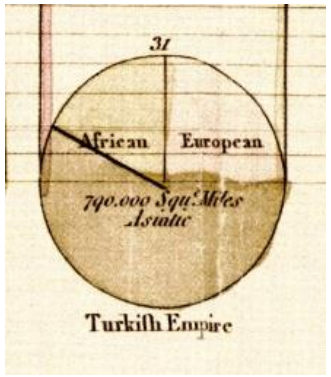
If there is time: Q-Q plots

A very very brief history of data visualization

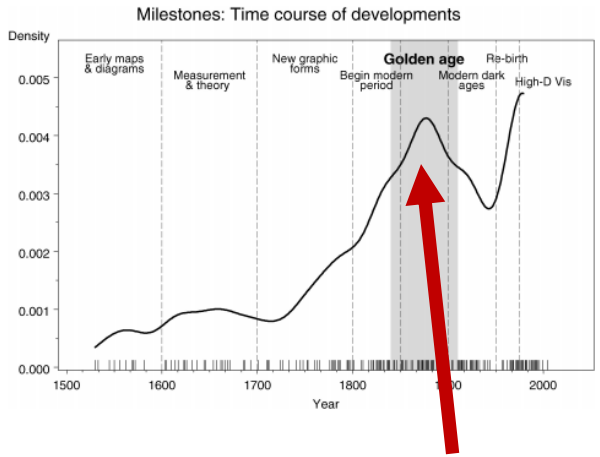
The age of modern statistical graphs began around the beginning of the 19th century

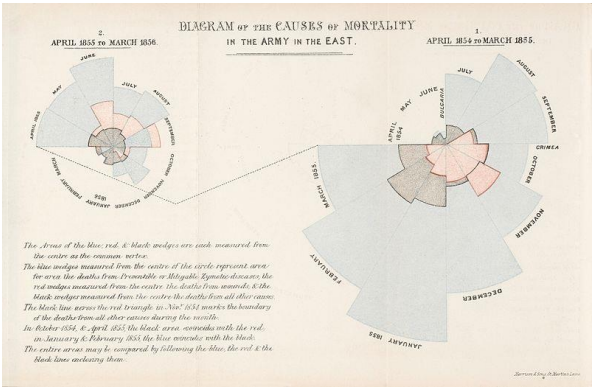
William Playfair (1759-1823)

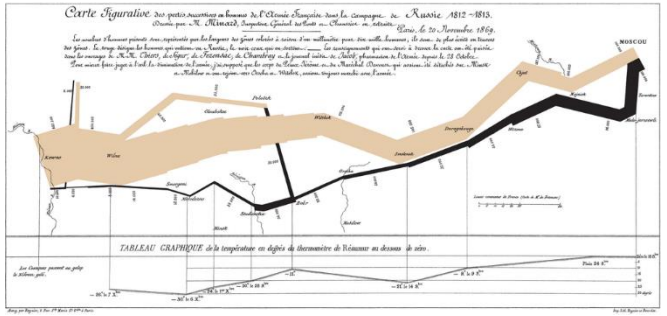


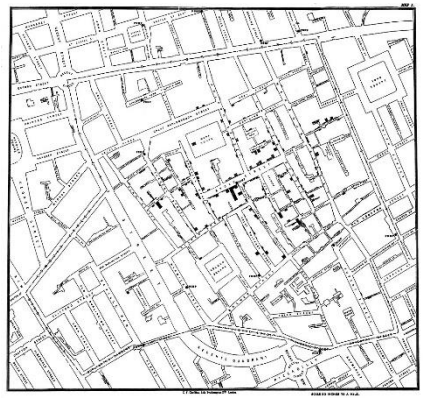


According to Friendly, statistical graphics researched its golden age between 1850-1900



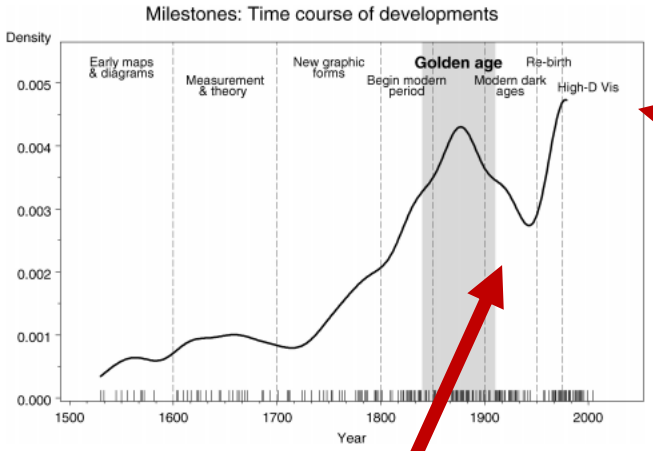


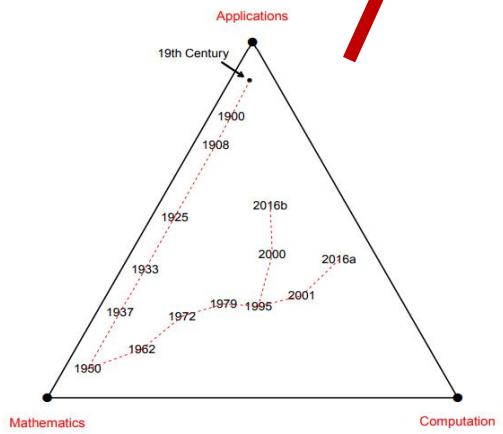


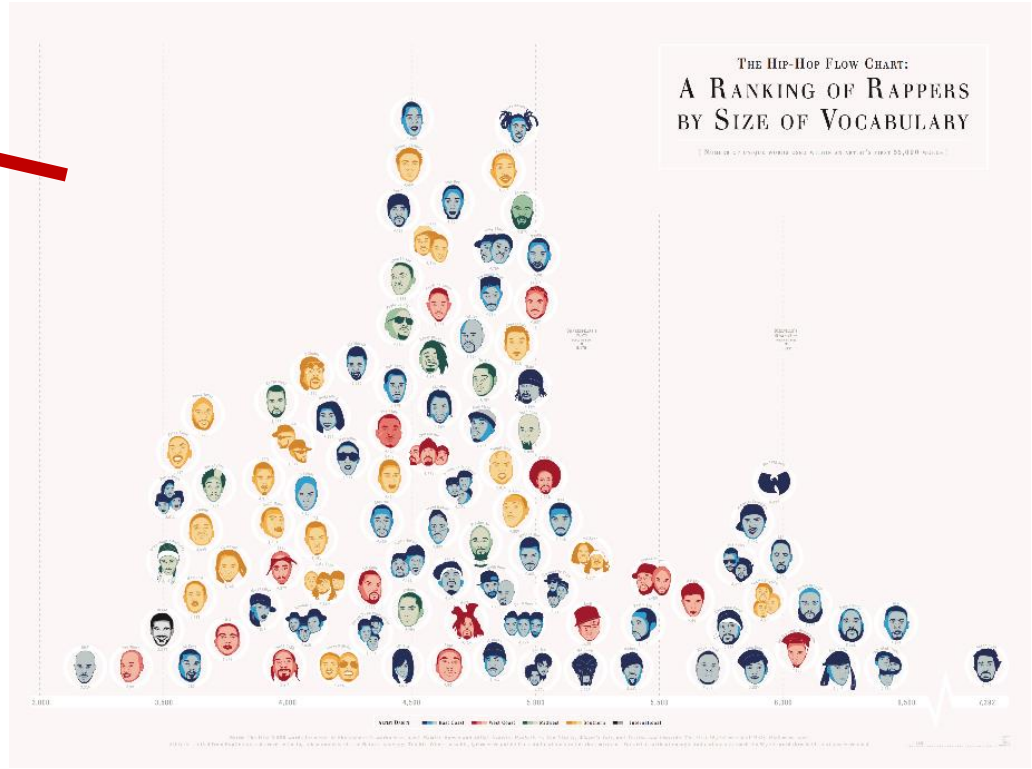


A very very brief history of data visualization

"Graphical dark ages" around 1950



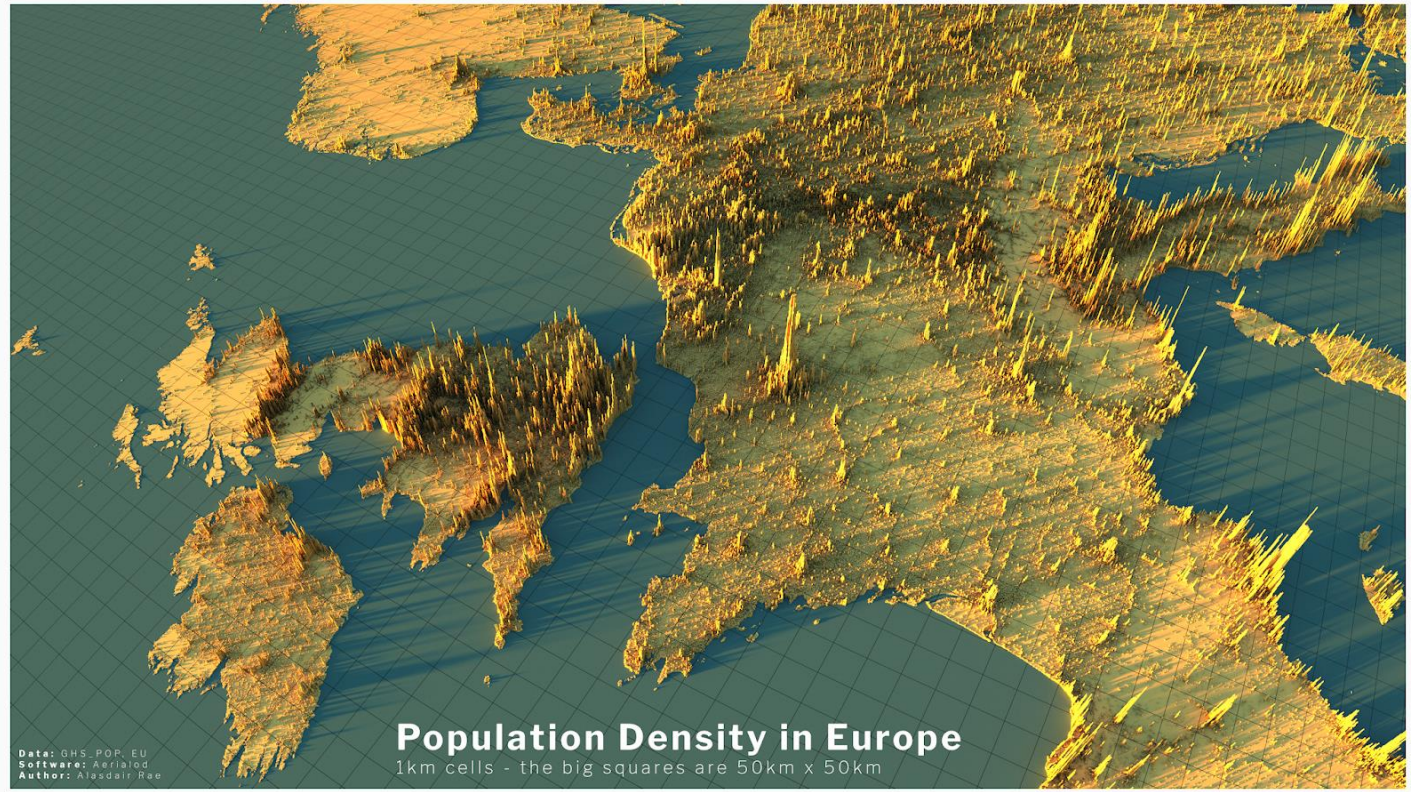


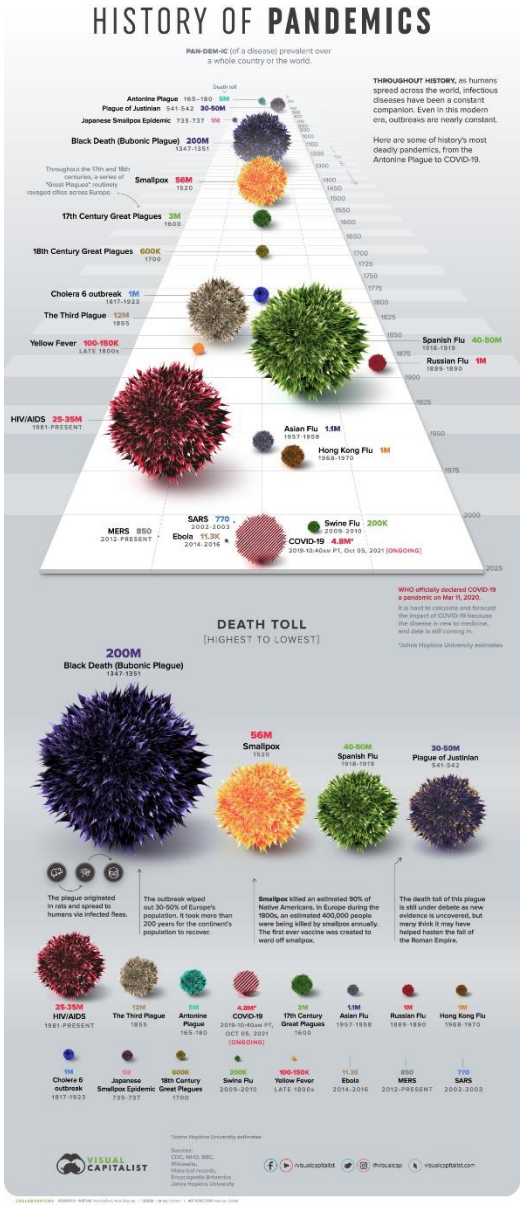


Currently undergoing a "Graphical re-birth"

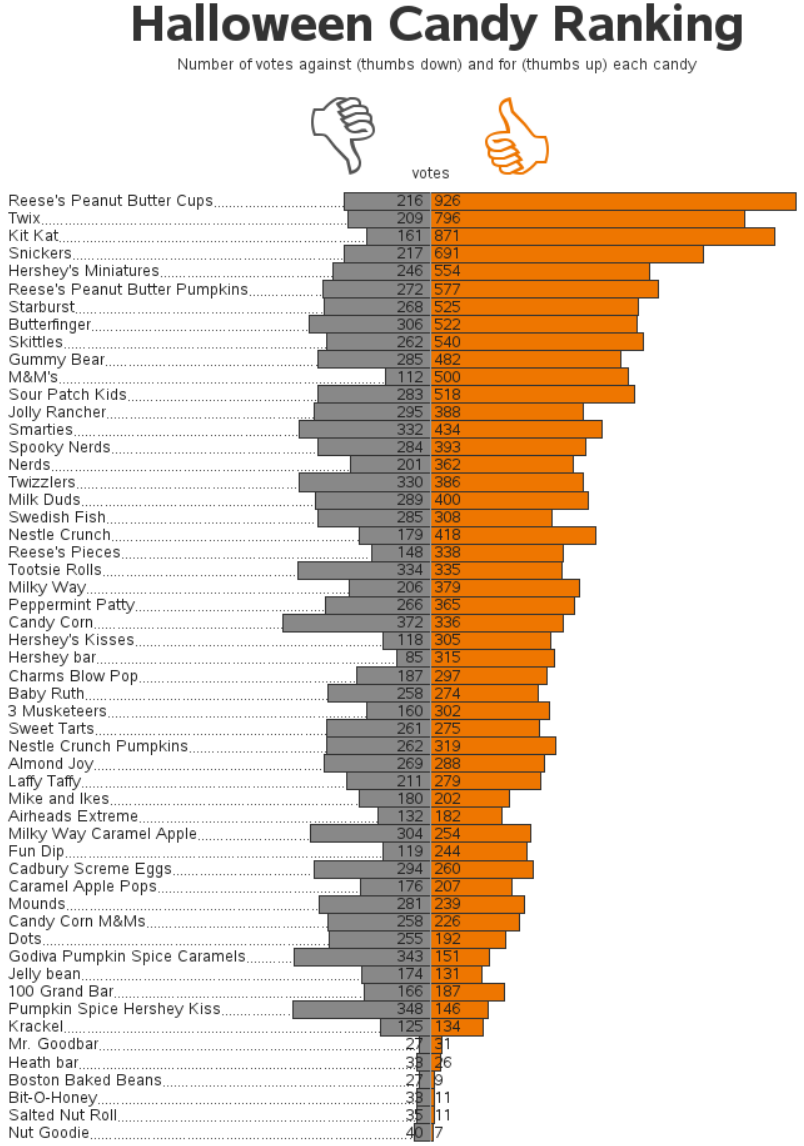
Computer Age Statistical Inference, Efron and Hastie

Your visualizations

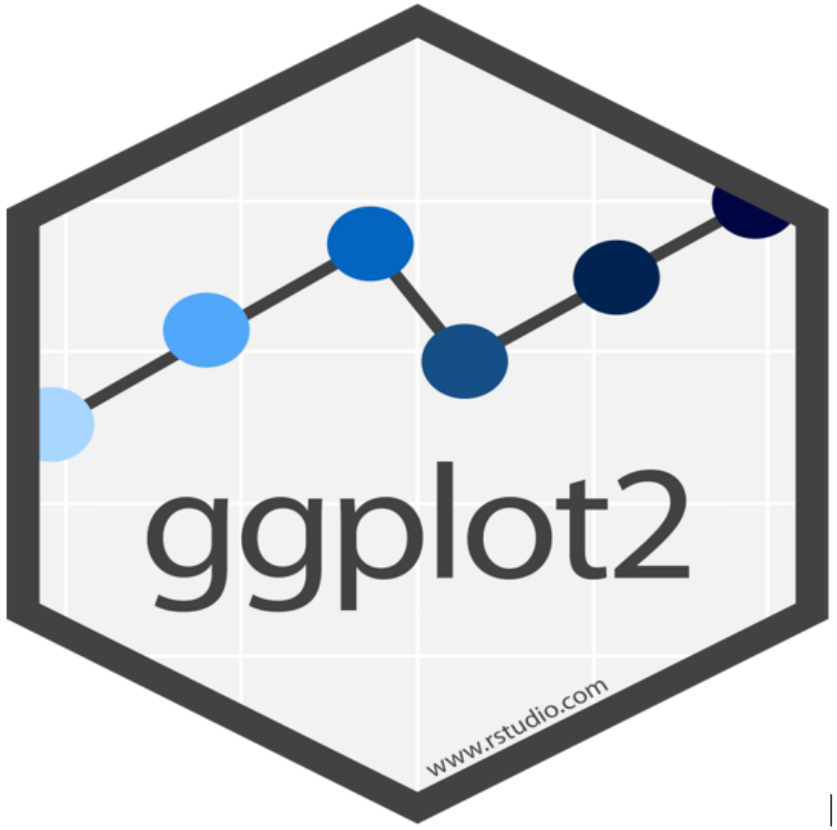




Your visualizations



Additional features of ggplot



Review/continuation of ggplot

A Frame: Coordinate system on which data is placed

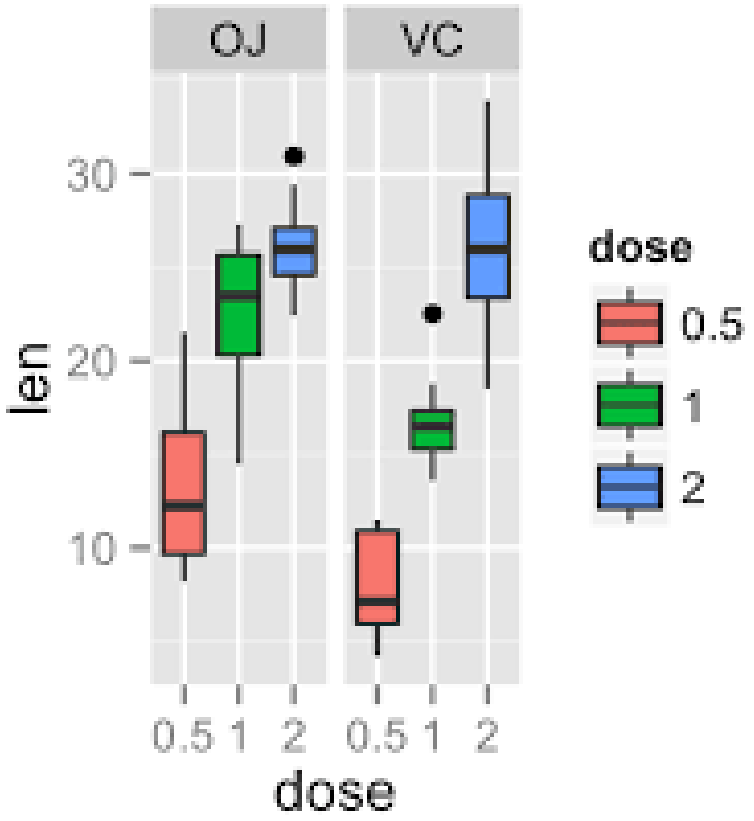
Glyphs: basic graphic unit representing cases or statistics

Scales and guides: shows how to interpret axes and other properties of the glyphs

Facets: allows for multiple side-by-side graphs based on a categorical variable

Layers: allows for more than one types of data to be mapped onto the same figure

Theme: contains finer points of display (e.g., font size, background color, etc.)



Review/continuation of ggplot

A Frame: Coordinate system on which data is placed

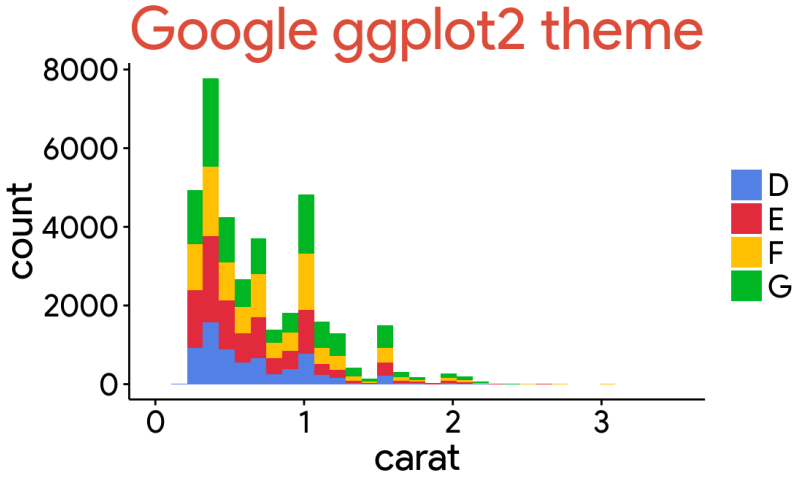
Glyphs: basic graphic unit representing cases or statistics

Scales and guides: shows how to interpret axes and other properties of the glyphs

Facets: allows for multiple side-by-side graphs based on a categorical variable

Layers: allows for more than one types of data to be mapped onto the same figure

Theme: contains finer points of display (e.g., font size, background color, etc.)



ggplot bonus features: emojis

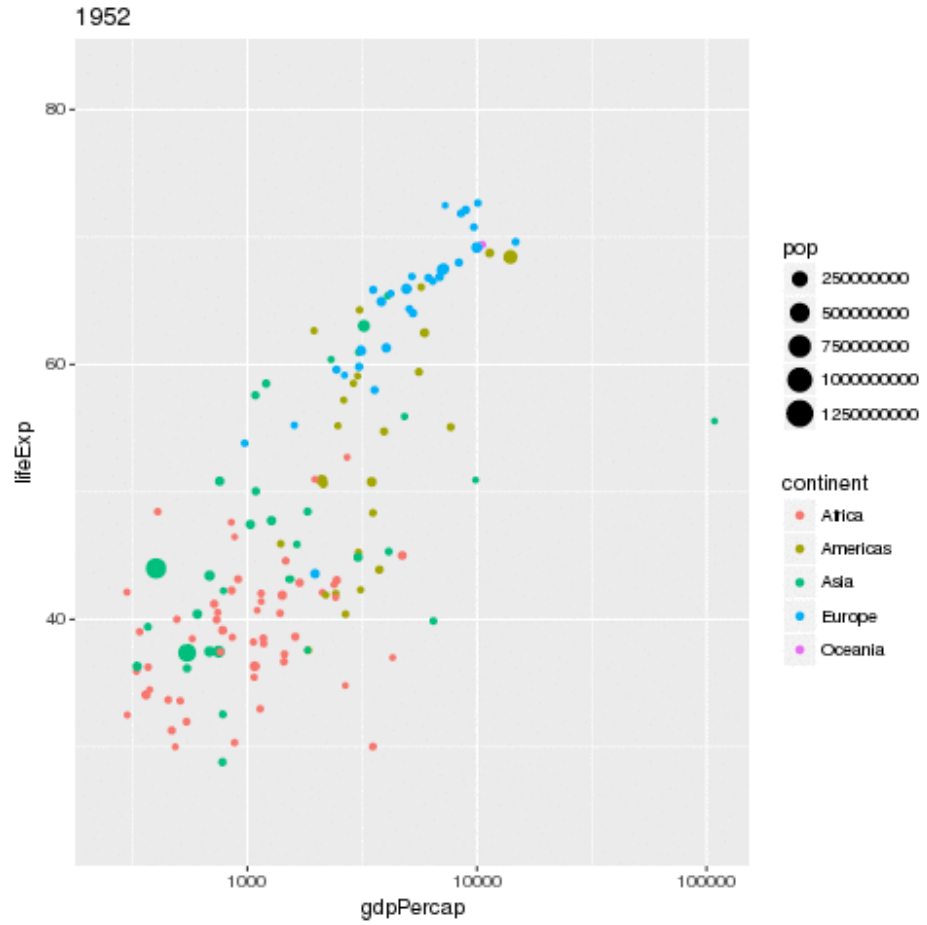
There are also additional packages that add more geoms

```
>  library(emoGG)

> ggplot(mtcars, aes(wt, mpg)) +
        geom_emoji(emoji="1f697")
```

ggplot bonus features: animation

We can create animated images (gifs) using the gganimate package

```
library(gganimate)

ggplot(gapminder, aes(gdpPercap, lifeExp,
          size = pop, col = continent)) +
  geom_point(alpha = 0.7, show.legend = FALSE) +
  scale_x_log10() +
 # Here comes the gganimate specific bits
   labs(title = 'Year: {frame_time}',
            x = 'GDP per capita', y = 'life expectancy') +
   transition_time(year) +
   ease_aes('linear')
```



Let's try it in R...

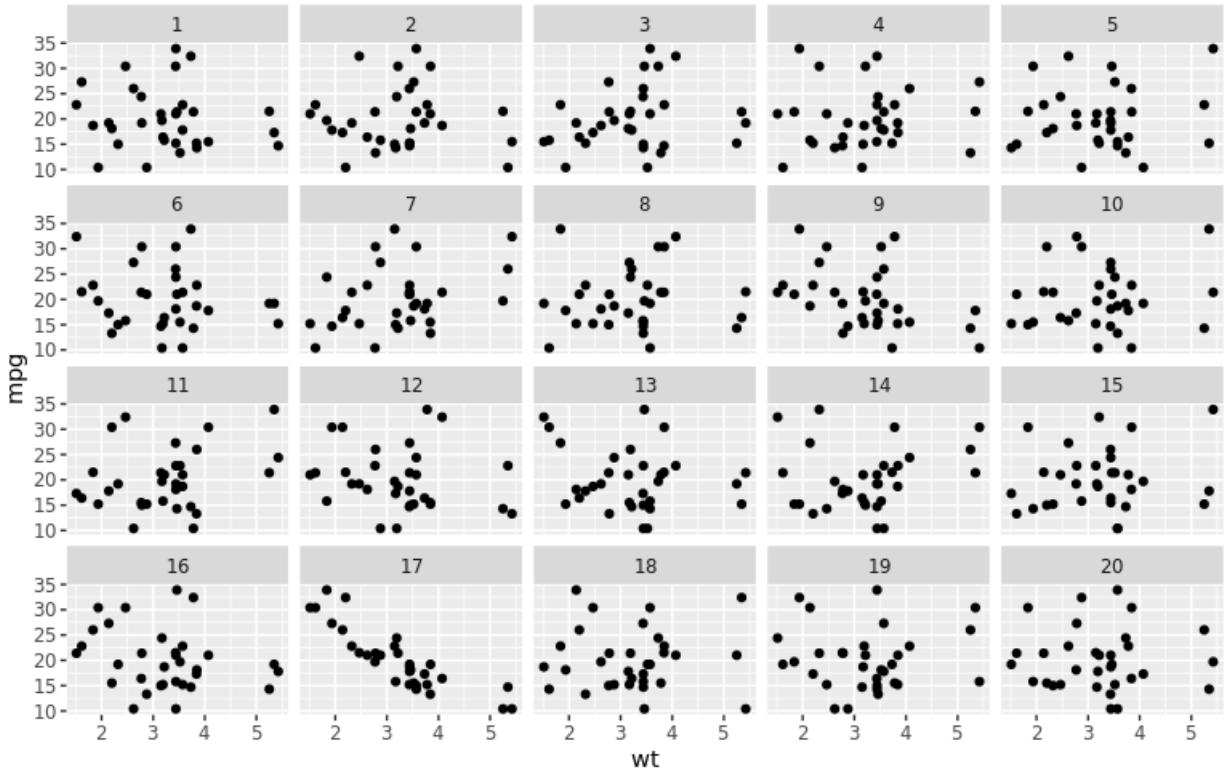
Visual hypothesis test

In visual hypothesis tests, we create data visualizations to try to assess whether particular relationships exist in our data.

- One way this is done through a visual lineup.

Visual hypothesis test

Which plot shows the true relationship between a car's weight and the number of miles per gallon a car gets?



Let’s try it in R…

Writing functions

We've used many R functions in this class

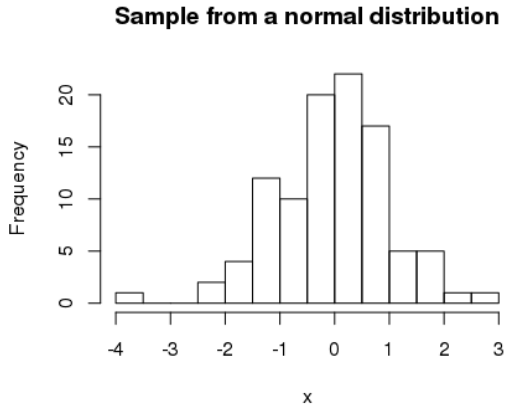
Let's explore writing our own functions!

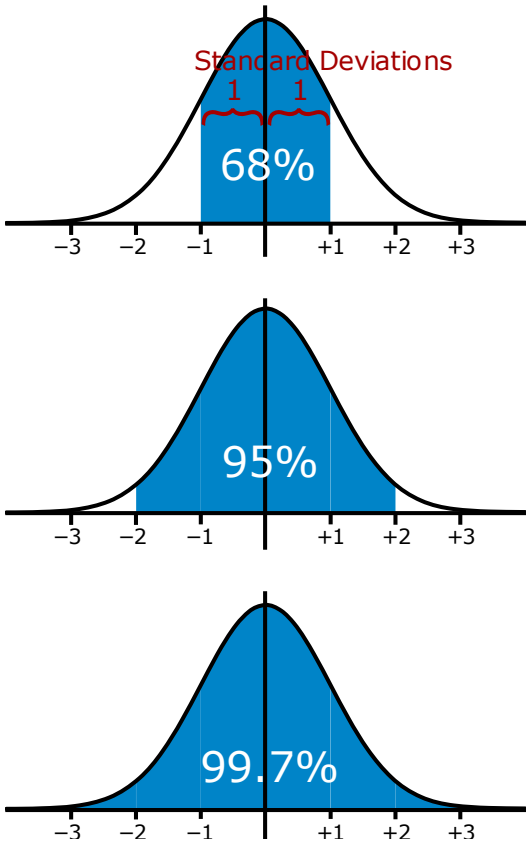
Quantile-Quantile plots

Density functions

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

A **density curves** are mathematical functions f(x) that are used to calculate probabilities






```
dnorm(x, 0, 1)
rand_data <- rnorm(100, 0, 1)
hist(rand_data)
```

How can you assess whether data comes from a particular distribution?

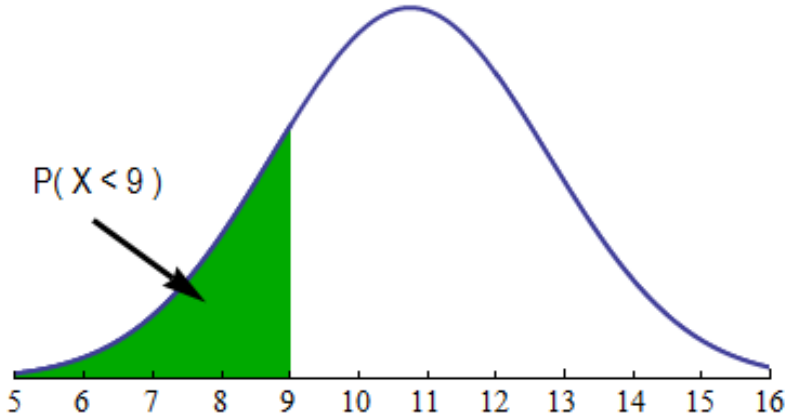
Quantiles

As you know, to get the probability (area) from a normal distribution we can use the pnorm function

```
pnorm(x, mu, sigma)
```

e.g., P(X < 9; μ = 11, σ = 3)

pnorm(9, 11, 3) = 0.2525



The quantile function is the inverse of the probability functions.

For a given probably *p*, (area between 0 and 1), it tells us the x value such that P(X < x) = *p*.

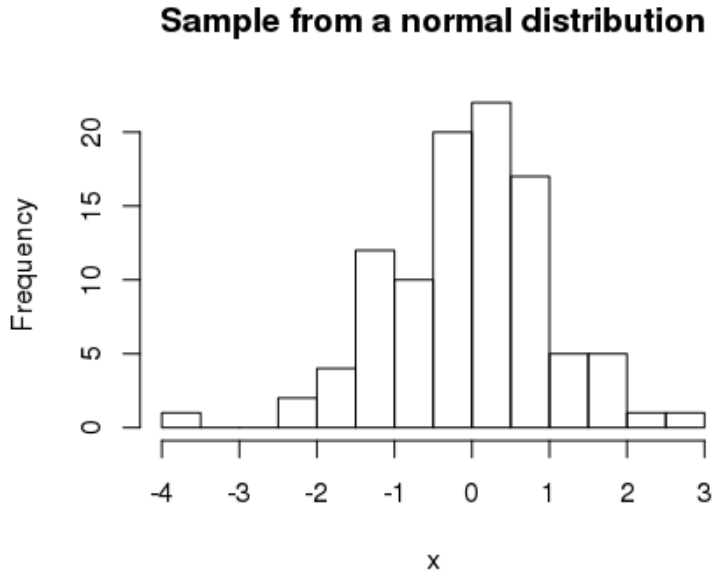
```
qnorm(p, mu, sigma)
```

e.g., P(X < ?; μ = 11, σ = 3) = 0.252

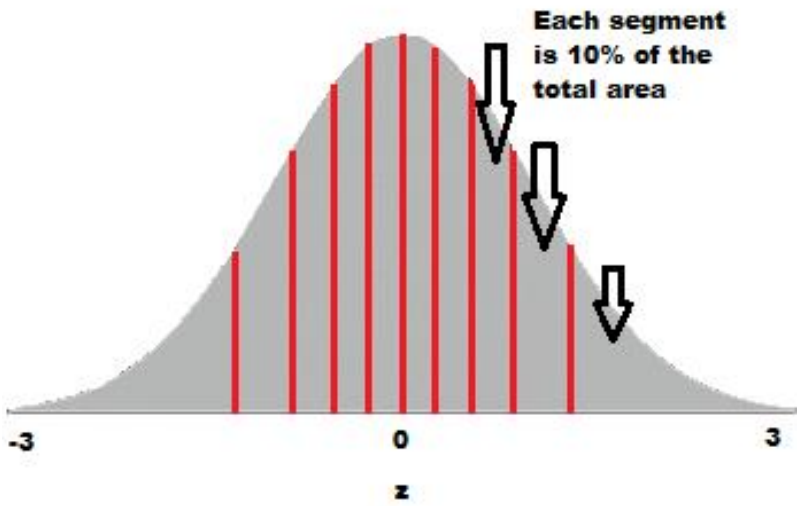
qnorm(.2525, 11, 3) = 9

Quantile-quantile plots (Q-Q plots)

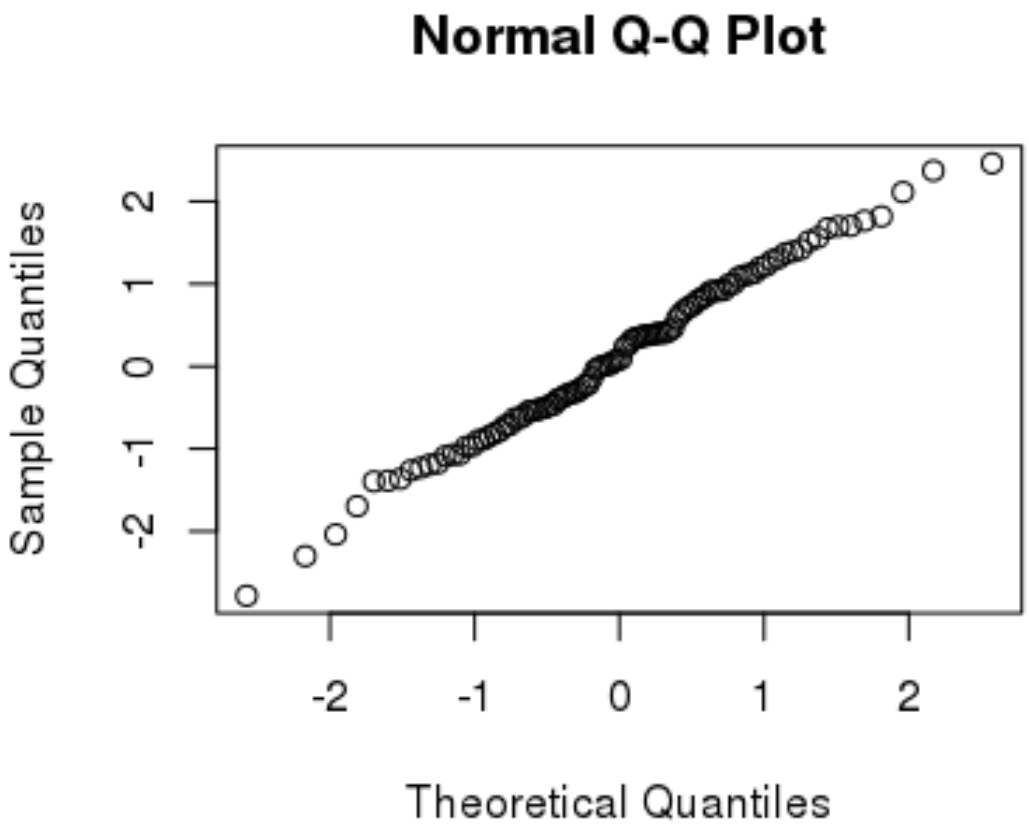
Quantile-quantile plots (Q-Q plots) can be used to assess whether a data sample comes from a particular distribution

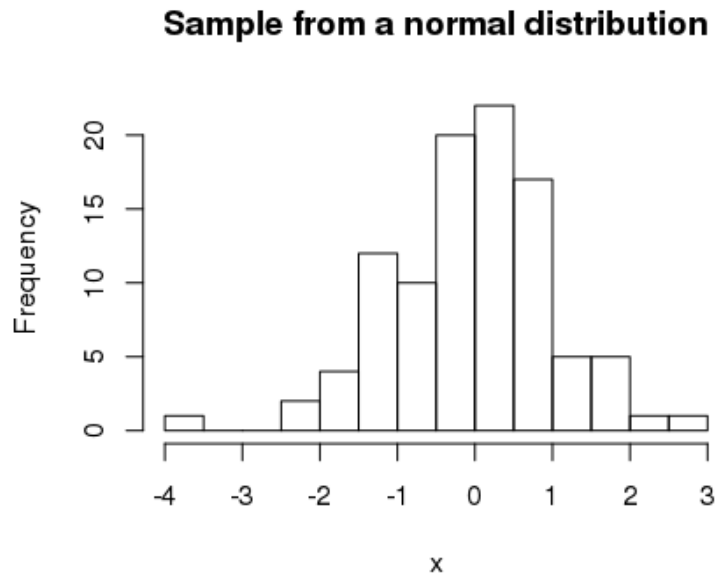


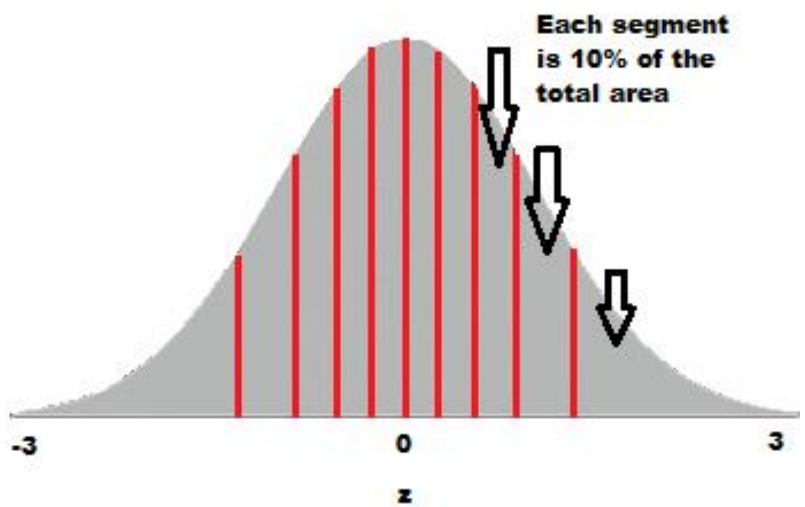
Q-Q plots show the observed quantile values from a data sample against the theoretical quantile values from a known distribution



Quantile-quantile plots (Q-Q plots)







Let’s try it in R…

Summary of R probability functions

Plot the actually density curve

- dnorm(x_vec, mu, sigma)

Get the probability that we would get a random value less than x

- pnorm(x_vec, mu, sigma)

Get the quantile value for a given proportion of the distribution

- qnorm(area, mu, sigma)

Note: pnorm and qnorm are inverses of each other

- y = pnorm(x, mu, sigma)
- qnorm(y, mu, sigma) # the output value here is x