

**ODDS**

**AND**

**ENDS**

# Overview

Discussion of data visualizations

Visual hypothesis test

Q-Q plots

If there is time: writing functions

- Writing functions
- Conditional statements
- Using computer simulations to assess confidence interval coverage

# Data visualization

What interesting data visualizations did you find?

Let's spend ~3 minutes discussing the interesting visualizations you found in groups of 3



# ggplot bonus features to try on your own: emojis

There are also additional packages that add more geoms

```
> library(emoGG)
```

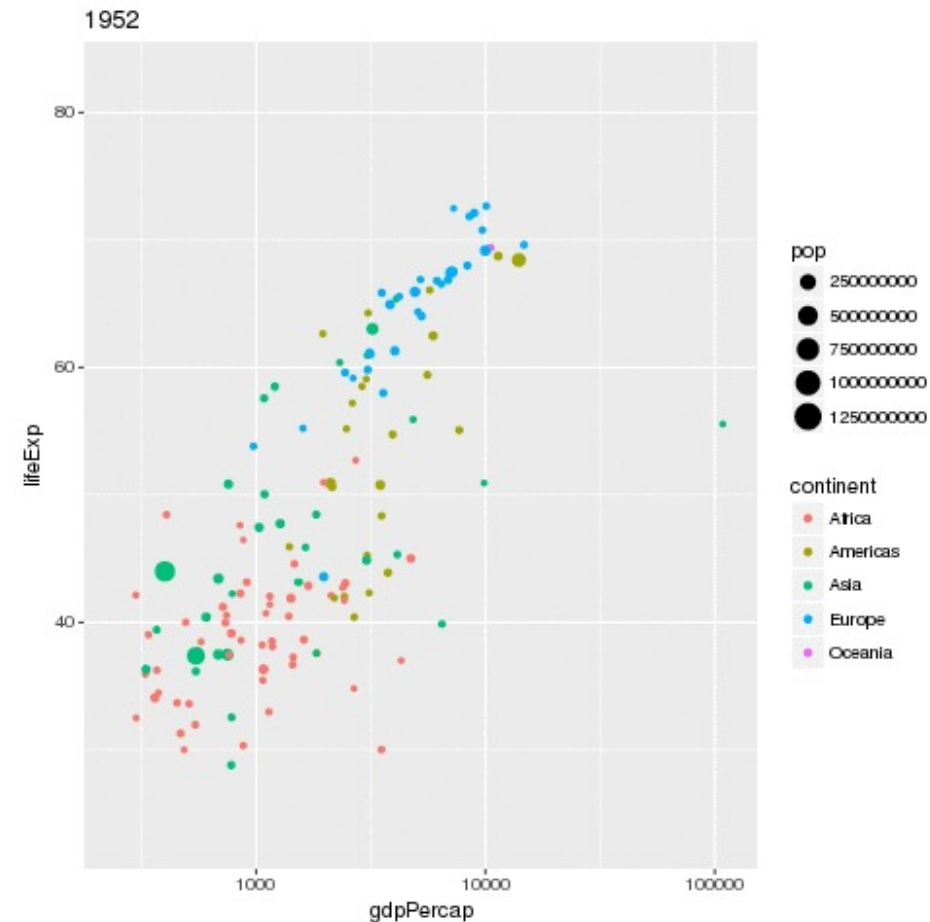
```
> ggplot(mtcars, aes(wt, mpg)) +  
  geom_emoji(emoji="1f697")
```

# ggplot bonus features to try on your own: 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')
```



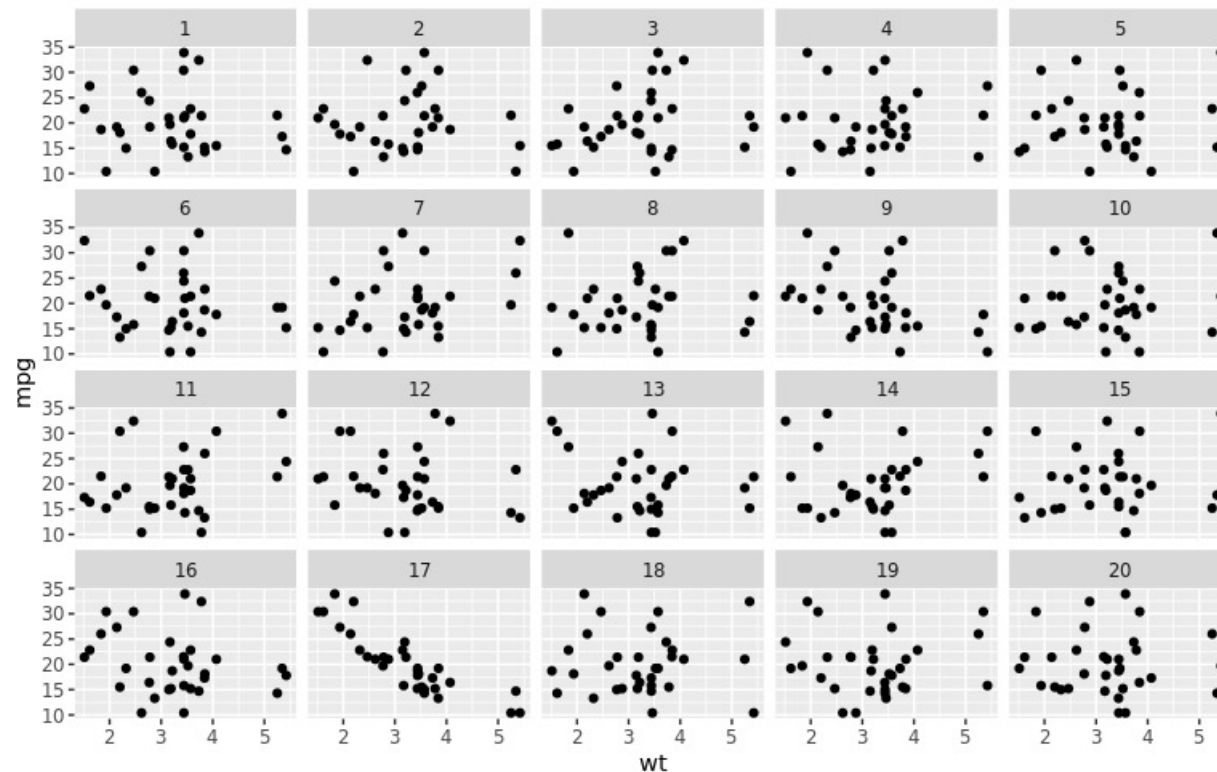
# Visual hypothesis test

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...

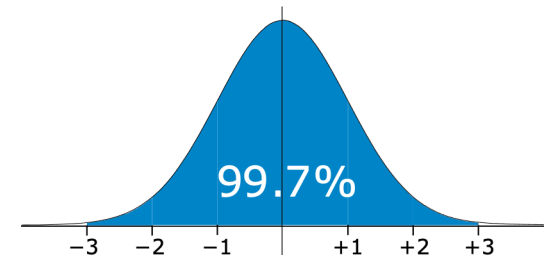
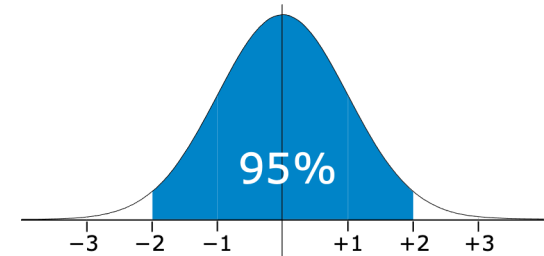
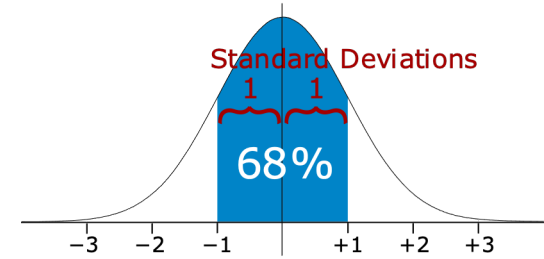
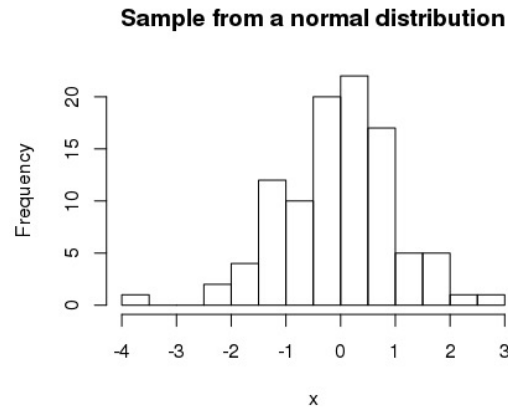


# Quantile-Quantile plots

# 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)
```

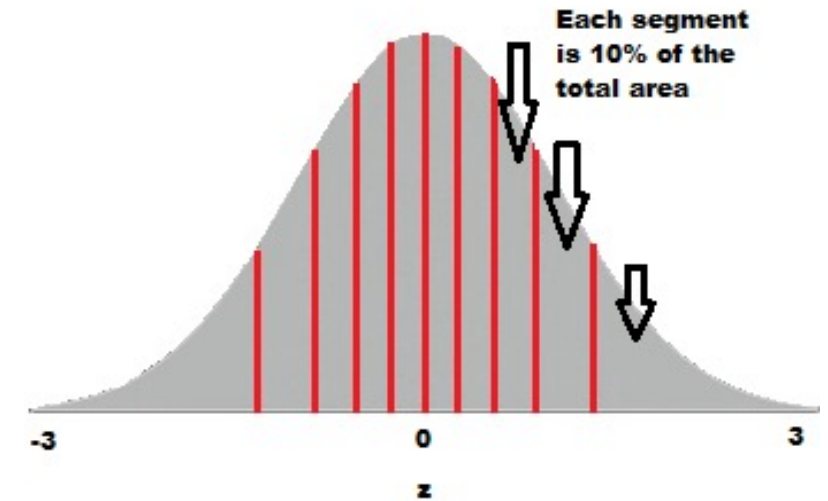
```
hist(rand_data)
```

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

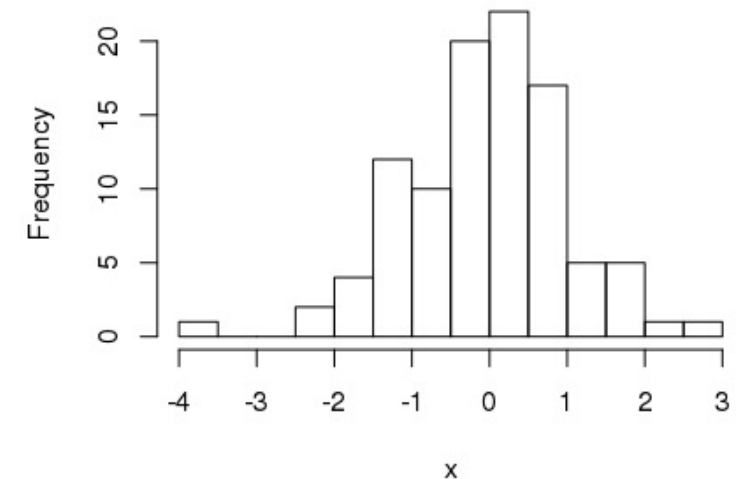
# Quantile-quantile plots (Q-Q plots)

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

They plot the observed quantile values from a data sample against the theoretical quantile values from a known distribution

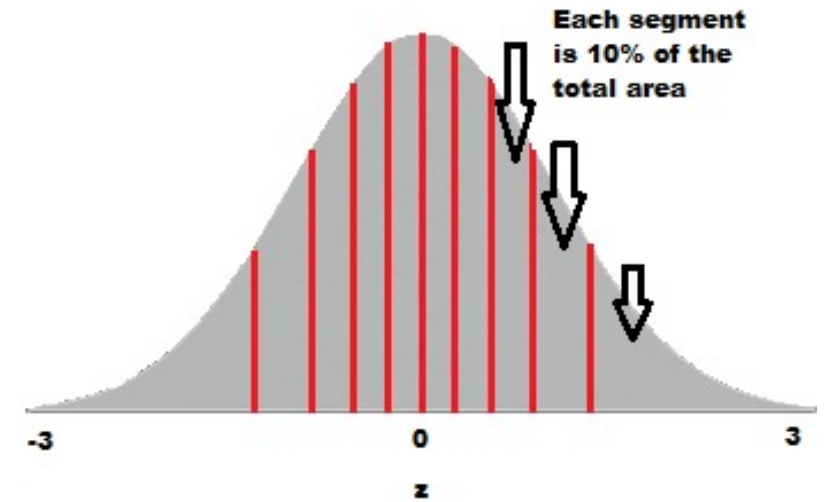
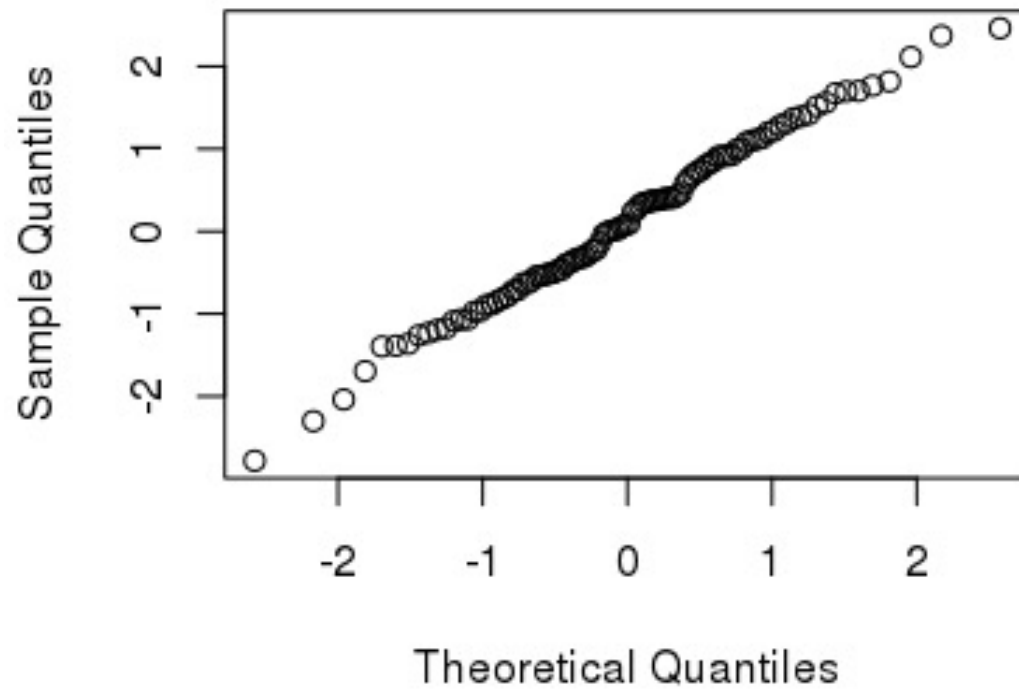


Sample from a normal distribution

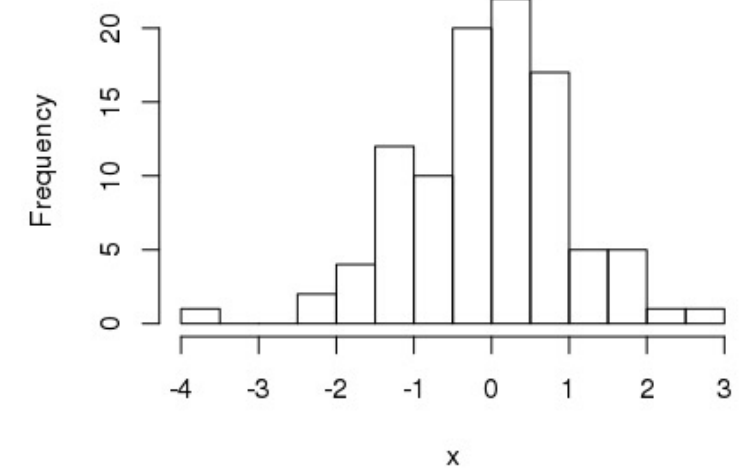


# Quantile-quantile plots (Q-Q plots)

**Normal Q-Q Plot**



**Sample from a normal distribution**



Let's try it in R...

# Writing functions

We've used many R functions in this class

Let's explore writing our own functions!