Lecture 1
Segment 2

Write a script

Write a script

- Script
  - Several lines of code
  - Composed in the R editor
  - Comment and save for later use

Write a script

- Goals for today's script
  - Read data into a dataframe
  - Explore the contents of the dataframe
  - Plot histograms
  - Get descriptive statistics

Write a script

- Example: Wine tasting!
  - Data are available in the following file:
    - STATS1.EX.01.TXT

Write a script

- First line(s) of code should be comments

```
# Statistics One, Lecture 3, example script
# Read data, plot histograms, get descriptives
```

Write a script

- Read data into a dataframe called “ratings”

  ratings <- read.table(“STATS1.EX.01.TXT”, header = T)

Write a script

- Explore the contents of the dataframe

  class(ratings)

    R will return “data.frame”

Write a script

- Explore the contents of the dataframe

  names(ratings)

    R returns “RedTruck WoopWoop HobNob FourPlay”

Write a script

- Plot histograms

  hist(RedTruck)

    R returns a histogram for RedTruck

Write a script

- Plot four histograms on one page

```
layout(matrix(c(1,2,3,4), 2, 2, byrow = TRUE))
hist(ratings$WoopWoop, xlab = "Rating")
hist(ratings$RedTruck, xlab = "Rating")
hist(ratings$HobNob, xlab = "Rating")
hist(ratings$FourPlay, xlab = "Rating")
```

Write a script

- Get descriptive statistics

  describe(ratings)

Write a script

- Functions used:
  - read.table
  - class
  - names
  - hist
  - describe

Learn more about functions

- In R console, type:

  >help(hist)

  >help(read.table)

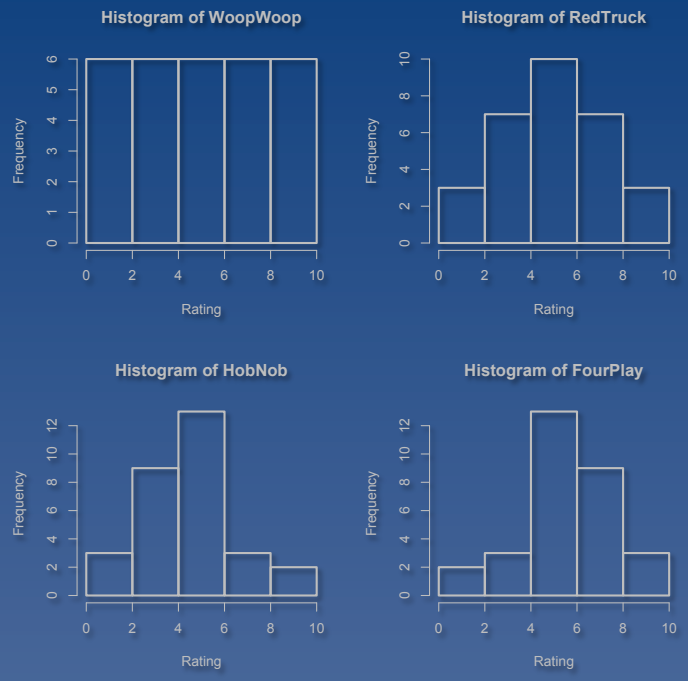
Final products

- Script
- Histograms
- Descriptive statistics

Script

```
# Statistics One, Lecture 3, example script
# Read data, plot historgrams, get descriptives
library(psych)

# Read the data into a dataframe called ratings
ratings <- read.table("STATS1.EX.01.TXT", header = T)

# What type of object is ratings?
class(ratings)

# List the names of the variables in the dataframe called ratings
names(ratings)

# Print 4 histograms on one page
layout(matrix(c(1,2,3,4), 2, 2, byrow = TRUE))

# Plot histograms
hist(ratings$WoopWoop, xlab = "Rating")
hist(ratings$RedTruck, xlab = "Rating")
hist(ratings$HobNob, xlab = "Rating")
hist(ratings$FourPlay, xlab = "Rating")

# Descriptive statistics for the variables in the dataframe called ratings
describe(ratings)
```

Histograms


Histogram of WoopWoop
Frequency
0 1 2 3 4 5 6
0 2 4 6 8 10
Rating
Histogram of RedTruck
Frequency
0 2 4 6 8 10
0 2 4 6 8 10
Rating
Histogram of HobNob
Frequency
0 2 4 6 8 10 12
0 2 4 6 8 10
Rating
Histogram of FourPlay
Frequency
0 2 4 6 8 10 12
0 2 4 6 8 10
Rating

Descriptive statistics

```
> # Descriptive statistics for the variables in the dataframe called ratings
> describe(ratings)
          var  n mean   sd median trimmed  mad min max range  skew kurtosis   se
RedTruck    1 30 5.50 2.26    5.5    5.50 2.22   1  10     9  0.00    -0.81 0.41
WoopWoop    2 30 5.50 2.92    5.5    5.50 3.71   1  10     9  0.00    -1.34 0.53
HobNob      3 30 5.03 2.01    5.0    4.96 1.48   1  10     9  0.33    -0.01 0.37
FourPlay    4 30 5.97 2.01    6.0    6.04 1.48   1  10     9 -0.33    -0.01 0.37
>
```

© 2012 Andrew Conway