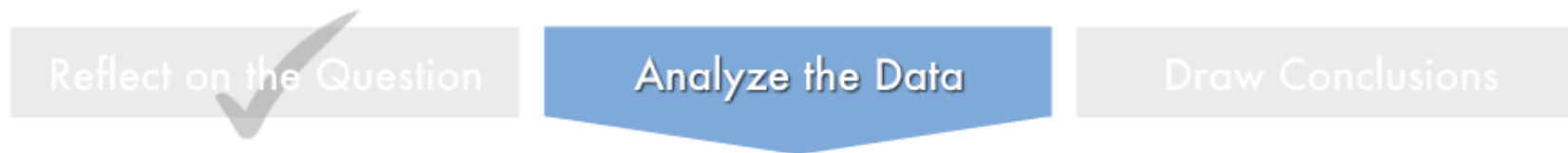**Courseware**   **Course Info**   **Discussion**   **Syllabus**   **Download R and RStudio**   **R Tutorials**   **Readings**   **Contact Us**

Progress   **Office Hours**   **Community**

Reflect on the Question ✓   Analyze the Data   Draw Conclusions

Primary Research Question

Is there a linear relationship between how often a rider places in the Top 10 and how often he stays on his bull?

Breakdown Your Analysis

Let's break this analysis into its required steps:

1. Create a scatterplot of the two variables of interest.
2. Check to see that the relationship is linear. Plot a line of best fit as a guide.
3. If the relationship is linear, calculate the correlation coefficient.
4. Carefully examine any outliers.
5. Interpret what the correlation says about the linear relationship between these variables.

Here is the code you will use:

```
# Visualize and describe the first variable of interest
hist(bull$RidePer)
fivenum(bull$RidePer)
mean(bull$RidePer)
sd(bull$RidePer)

# Visualize and describe the second variable of interest
hist(bull$Top10)
fivenum(bull$Top10)
mean(bull$Top10)
sd(bull$Top10)

# Create a scatterplot
plot(bull$RidePer,bull$Top10)

# Add line of best fit
abline(lm(bull$Top10~bull$RidePer))

# Calculate the correlation coefficient
cor(bull$RidePer,bull$Top10)

# Create a correlation matrix
vars <- c("Top10", "RidePer")
cor(bull[,vars])
```

Help

(1/1 point)

What will this line of the code do for us?

hist(bull$RidePer)

Help

- ○ It will pull out data on just the bulls (but not the riders) from the BullRiders datafile.
- ● It will generate a histogram of the successful ride percentages of all the riders.   ✔
- ○ It will read the first few cases in the dataset and produce them in the console window.

Correct. The "hist" function will produce a histogram for the specified variable, which in this case is RidePer, or the percentage of successful rides for each rider.

| Final Check | Save | Hide Answer | *You have used 1 of 2 submissions*

(1/1 point)

What will appear in the scatterplot produced by this line of code?

plot(bull$RidePer,bull$Top10)

- ● There will be a single data point for each bull rider.   ✔
- ○ There will be two data points for each rider (one for each variable).

Correct. The scatterplot will plot the value of "Top10" on "RidePer" for each bullrider.

| Final Check | Save | Hide Answer | *You have used 1 of 2 submissions* |

(1/1 point)

Which value is **not** a possible output of the following line of code?

```
cor(bull$Rides,bull$Top10)
```

- ○ 0.75
- ● -1.02 ✔
- ○ 0.61
- ○ -0.04

Correct. -1.02 is not a possible Pearson correlation coefficient value. Correlation values must be equal to or between -1 (a perfect negative linear relationship) and 1 (a perfect positive linear relationship). The absolute value represents the magnitude of the correlation, while the sign of the value represents the direction of the correlation.

| Final Check | Save | Hide Answer | *You have used 1 of 2 submissions* |

(1/1 point)

A correlation matrix allows you to calculate multiple correlation coefficients at a time. Here, we are only asking for the

correlation between RidePer and Top10. If you wanted to include other variables as well, how would you do that?

```
# Create a correlation matrix
vars <-c("Top10", "RidePer")
cor(bull[,vars])
```

○ Eliminate the line cor(bull[,vars]), because a correlation is only between two variables.

● Add the variable names to the "vars" object.   ✔

○ You couldn't do that. You would have to run separate correlation matrices for each pair of variables.

That's correct. By including additional variables into the "vars" object, we can create a correlation matrix for as many variables in our dataset as we wish.

| Final Check | Save | Hide Answer | *You have used 1 of 2 submissions*

About edX   Follow Us

EdX offers interactive online classes and MOOCs from the world's best universities. Online courses from MITx, HarvardX, BerkeleyX, UTx and many other universities. Topics include biology, business, chemistry, computer science, economics, finance, electronics, engineering, food and nutrition, history, humanities, law, literature, math, medicine, music, philosophy, physics, science, statistics and more. EdX is a non-profit online initiative created by founding partners Harvard and MIT.

Terms of Service and Honor Code

Privacy Policy (Revised 4/16/2014)

Help

About

News

Contact

FAQ

edX Blog

Donate to edX

Jobs at edX

Twitter

Facebook

Meetup

LinkedIn

Google+