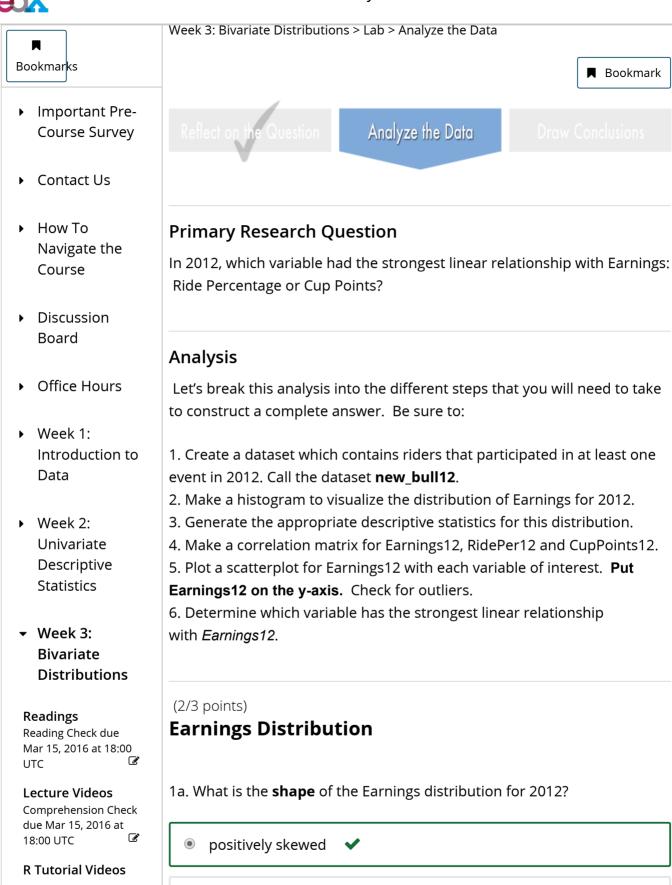


Pre-Lab

Lab

Pre-Lab due Mar 15, 2016 at 18:00 UTC

## UTAustinX: UT.7.10x Foundations of Data Analysis - Part 1



negatively skewed

Lab due Mar 15, 2016 at 18:00 UTC

## **Problem Set**

Problem Set due Mar 15, 2016 at 18:00 UT

- Week 4:
   Bivariate
   Distributions
   (Categorical
   Data)
- Week 5: Linear Functions

1b. What was the **average** amount earned by a bull rider? (Choose the appropriate measure of center; report without a \$ sign and round to the nearest whole number.)

111148 **X** Answer: 147952

1c. What was the **highest** amount earned by a bull rider? (Report without a \$ sign and round to the nearest whole number.)

You have used 1 of 1 submissions

(2 points possible)

## Make a Scatterplot of Earnings and Ride Percentage

2a. Does the scatterplot show a **linear** relationship?

No ▼ **X** Answer: Yes

2b. What is the **correlation** of Earnings with Ride Percentage for 2012? (round to three decimal places)

0.619 **X** Answer: 0.593

You have used 1 of 1 submissions

(2/2 points)

## **Create a Scatterplot of Earnings and Cup Points**

3a. Does the scatterplot show a linear relationship?

	hat is the <b>correlation</b> of Earnings with Cup Points for 2012? (report ee decimal places)
0.65	✓ <b>Answer:</b> 0.657
0.65	7
You I	have used 1 of 1 submissions
	ooints) liers and Influential Points
Some this in # ide which	tlier can have a significant impact on the correlation coefficient.  times it is important to remove these points to examine the size of  npact. Run this code to <b>identify</b> the extreme data value in Earnings:  entify specific case  n (new_bull12\$Earnings12 ==  new_bull12\$Earnings12))
4a. Th 	e extreme earnings data point belonged to the rider that came in Place in 2012. (Please spell your answer; do not use numerals.)
first	✓ Answer: First
	here does this data point fall in the scatterplot? (Make sure that ngs12 is on the y-axis)
•	Above the line 🗸
	Below the line
0	On the line
impac	remove this data point from the dataset to assess what kind of att, if any, it had on our correlation analysis. Run this code:  set the data  clier <- new bull12[new bull12\$Earnings12 < 1000000

Analyze the Data | Lab | UT.7.10x Courseware | edX ,] Then **rerun** the correlation matrix and the scatterplots to see the difference. Make sure to use the new dataframe (nooutlier) that you just created. 4c. After removing the outlier, what was the **new correlation** of Earnings and Ride Percentage for 2012? (Round to three decimals) Answer: 0.804 0.804 0.804 4d. After removing the outlier, what was the **new correlation** of Earnings and Cup Points for 2012? (Round to three decimals) **Answer:** 0.893 0.893 0.893 4e. We would say that this data point was an **influential point** because it caused the underlying relationship to be non-linear. • inflated the relationship between Earnings and the other variables. made the earnings of the other bull riders look less impressive than they really were. masked the strength of the relationships between Earnings and the other variables

You have used 1 of 1 submissions

© All Rights Reserved



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

















