

Student Developed Shiny Apps for Teaching Statistics

Background

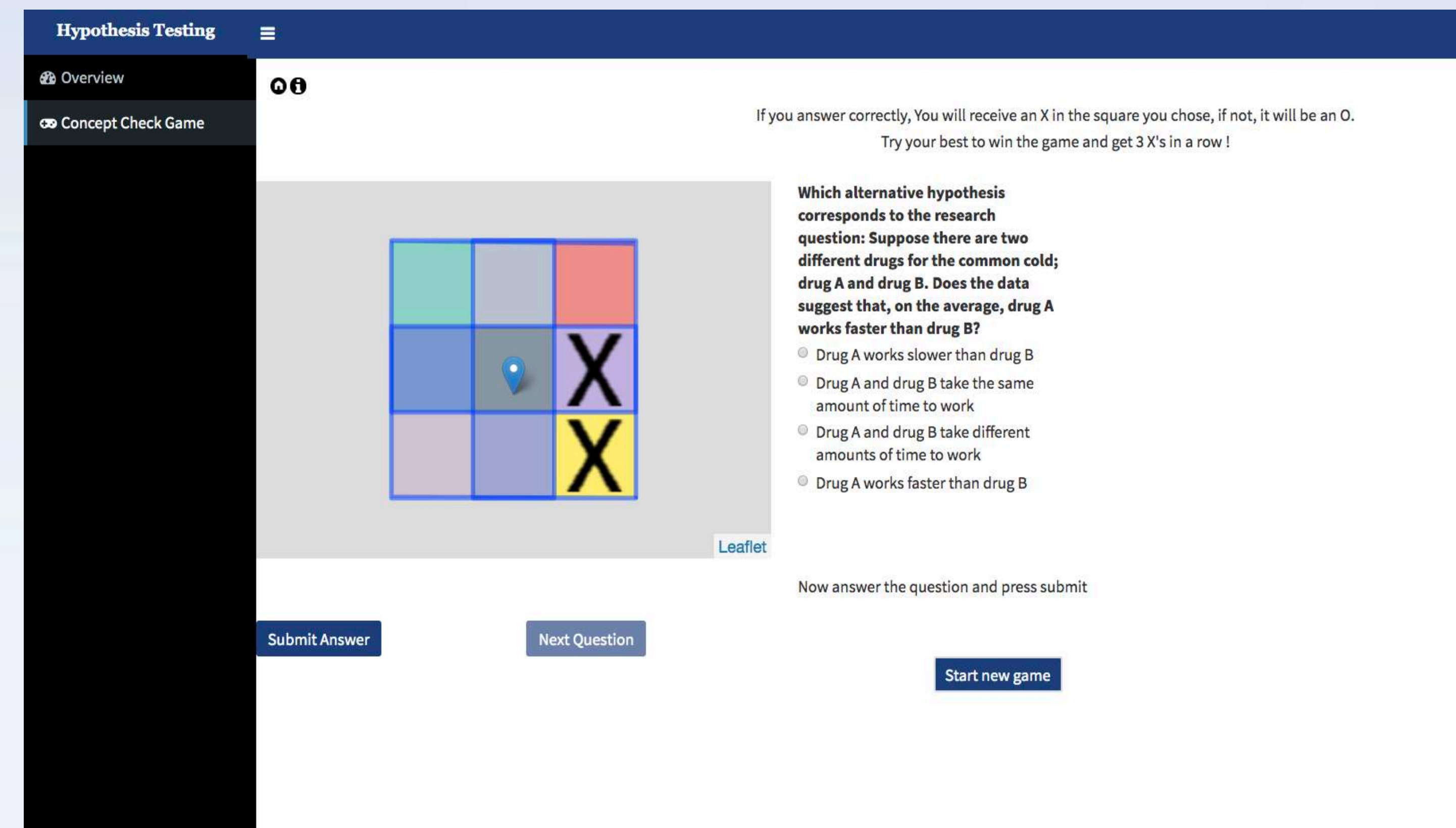
Web apps for teaching statistics can provide an active learning environment and dynamically illustrate concepts that are otherwise difficult to explain.

Many web apps have:

- Outdated Software (e.g. Flash, Java)
- Lack of consistency

R Shiny

- R Shiny is a package within R that helps to build web applications within R
- R Shiny can easily deal with both real-world data and quick interactive displays with user inputs
- Recent literature (see Fawcett, 2018 and Doi et al., 2016) show that implementing R Shiny apps can help students' confidence and overall course satisfaction.

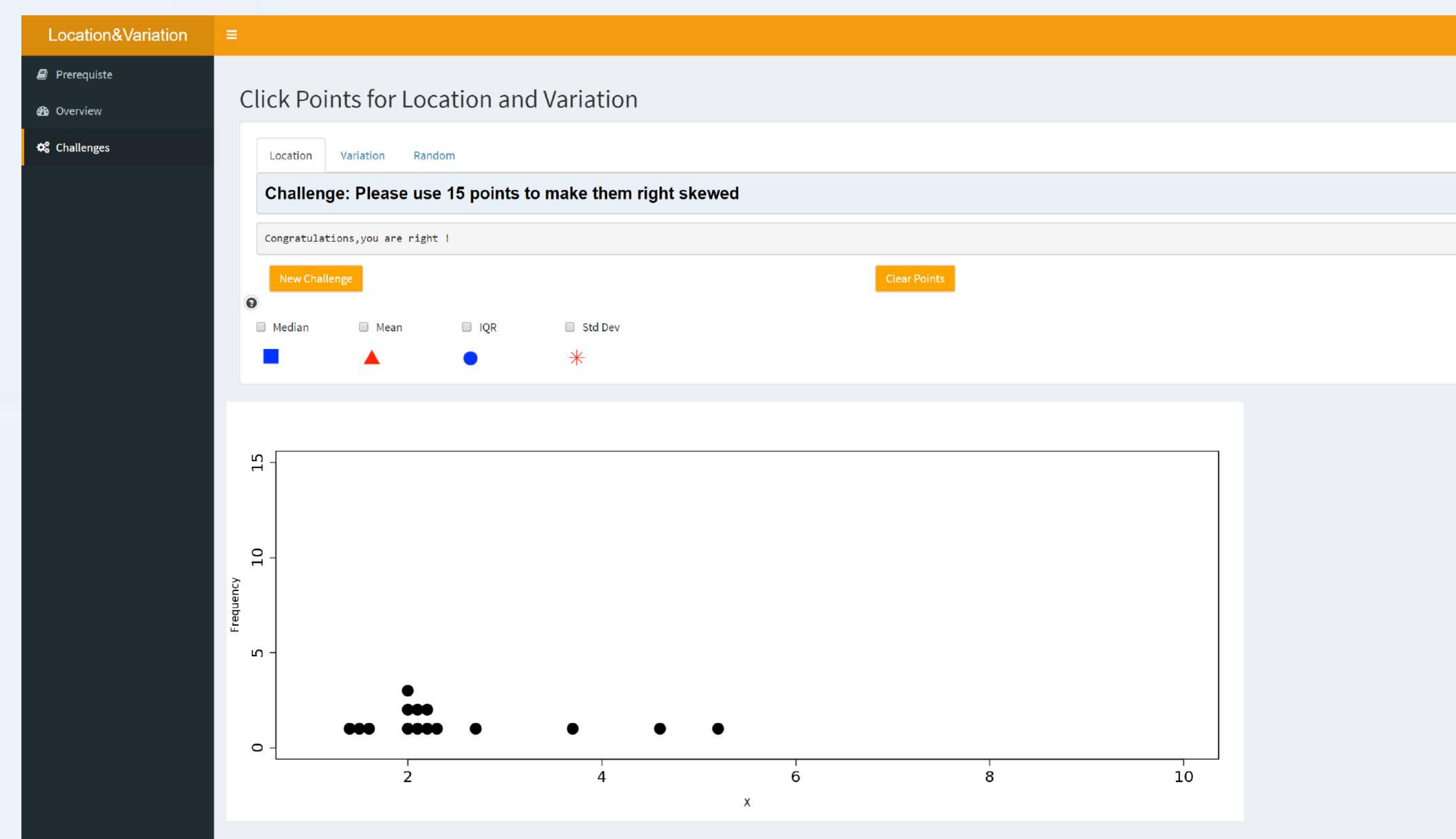


Lower division App Tested: Hypothesis Testing Game

Objective

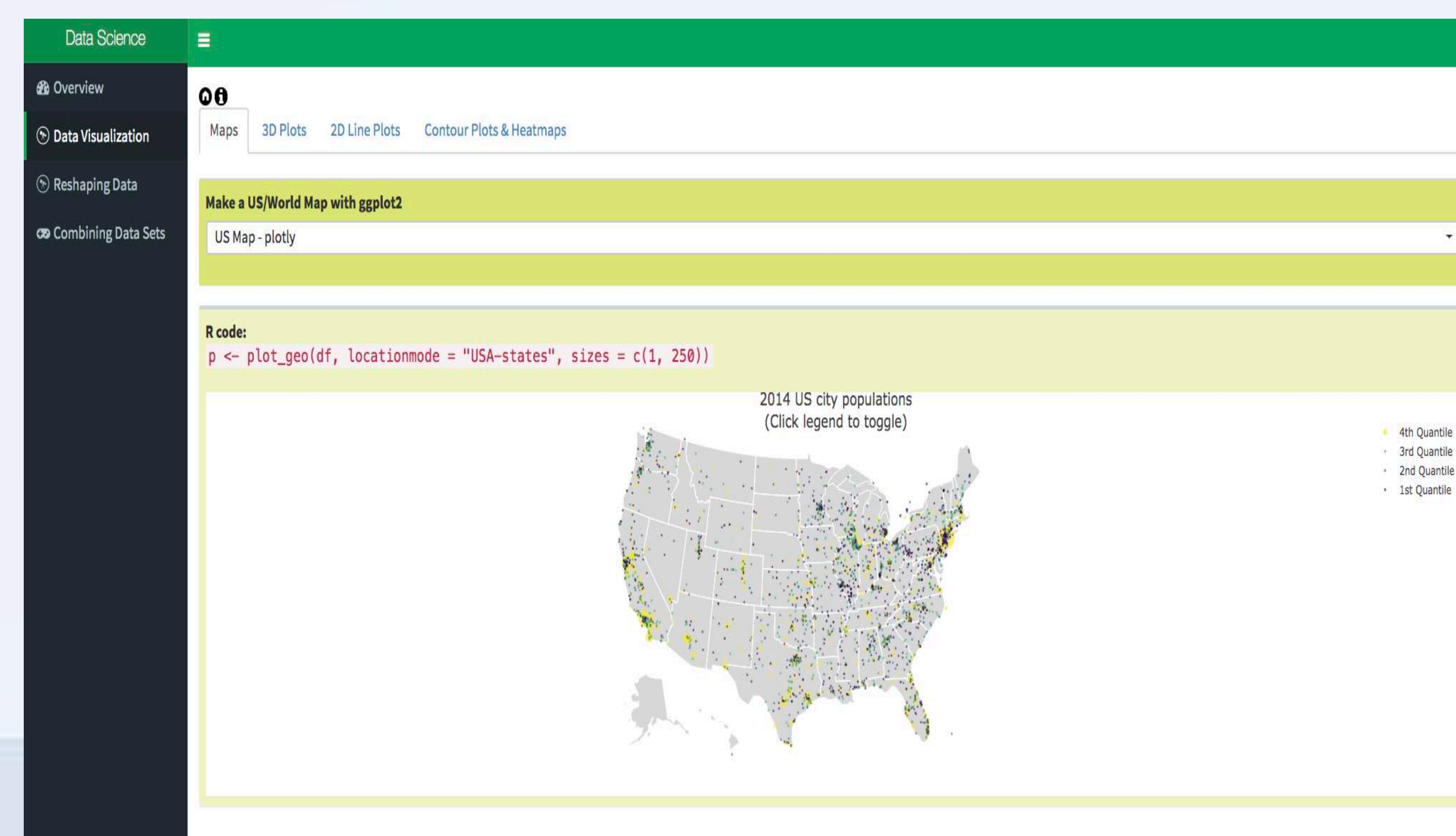
- To create a book of statistics apps covering multiple topics (Descriptive Statistics, Hypothesis Testing, Central Limit Theorem, ANOVA, Regression, Time Series, etc.) for teaching statistics classes through R Shiny applications
- To provide an authentic research opportunity to statistics majors serving as the app developers.
- To test the effectiveness of teaching applications using [the Book Of Apps for Statistics Teaching](#). The students in the study are recruited from those who enrolled in STAT 100 - Statistical Concepts and Reasoning at University Park campus.

Lower division App Tested: Location & Variation



Research Program

- R Shiny App Development: 10 undergraduate students per year in a 10 week summer program; five teams are formed with two people in each team, responsible for a specific topic
- The first half of the program focuses on debugging and improving the apps developed in the previous summer; the second half of the program focuses on developing new apps including upper-division statistical topics
- Students learned programming in R Shiny & GitHub
- App Effectiveness Testing: The IRB approved study was carried out in STAT 100: Introductory Statistics at Penn State University, University Park. Each week, one lab session used the student developed Shiny apps while the other lab session took a traditional lecturing approach.



Upper division App: Data Science

Method

- Create web-based apps including statistical concept explorations, real data investigations, and game formats
- The (non randomized) experimental study spanned six weeks during spring semester. During each week, one lab section used the Shiny applications for teaching, while the other section heard lectures and used clickers to respond to questions about the same topic (≈10 minutes of class time for each). Which section used the apps was alternated from week to week.
- A quiz was given at the end of each lab section. The scores on the quizzes will be compared between the treated and control conditions for each student.

Preliminary Results

Summary Data (individual student data not yet available)		% of students correct on Question 1		% of students correct on Question 2	
App	Section using app	With app	Without app	With app	Without app
Hypothesis Testing Caveats	afternoon	49/58 (84%)	31/54 (57%)	50/58 (86%)	46/54 (85%)
Hypothesis testing tic-tac-toe	morning	32/52 (62%)	29/61 (48%)	29/52 (56%)	25/61 (41%)
Confidence Intervals for proportions	afternoon	46/57 (81%)	32/54 (59%)	36/57 (63%)	33/54 (61%)
Central Limit Theorem	morning	Problems with website (incorrect link)			
Law of Large Numbers	afternoon	39/62 (63%)	20/52 (38%)	36/62 (58%)	16/52 (31%)
Location and Variation	morning	No data yet on individual app related questions			

Acknowledgment

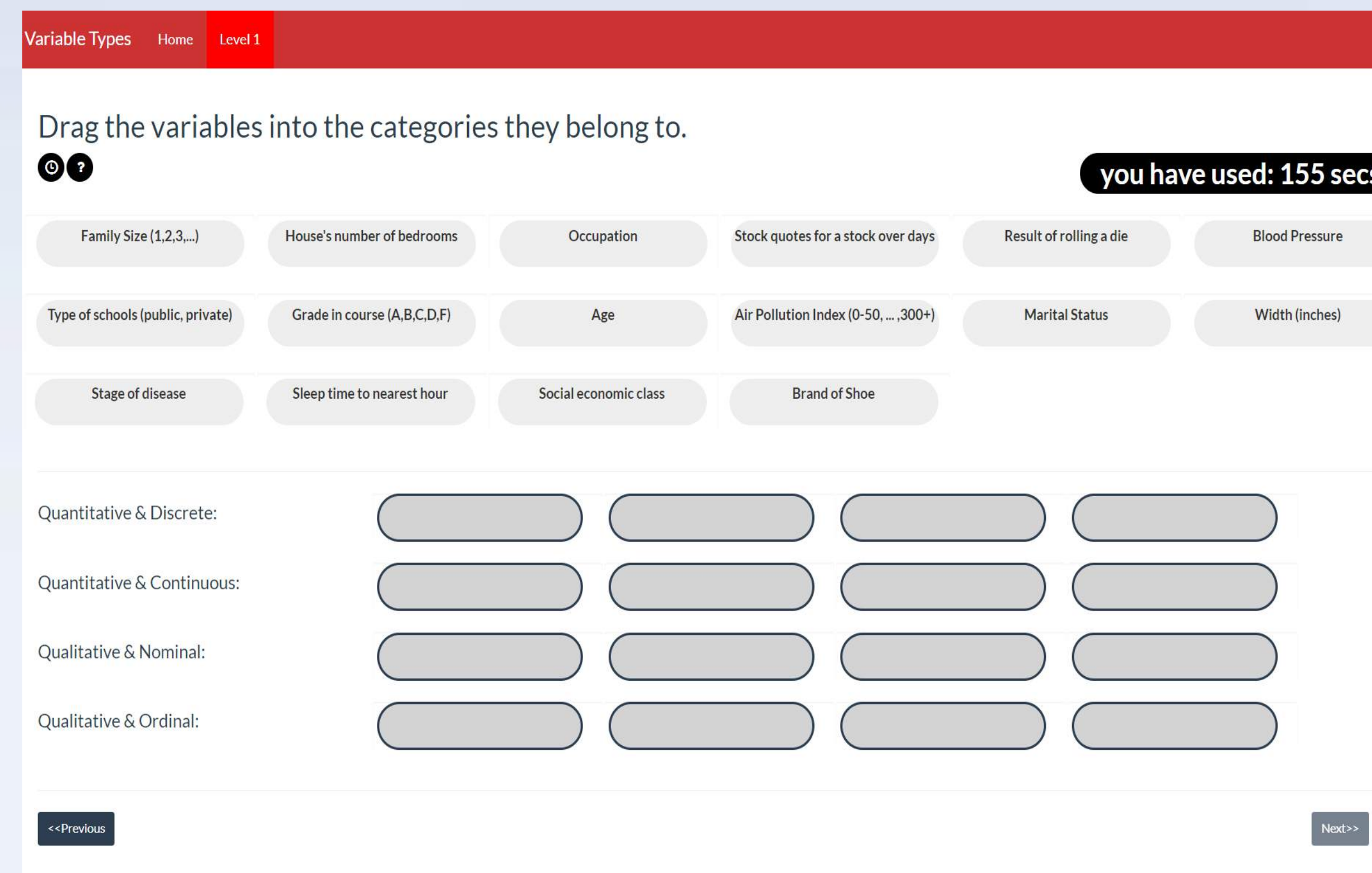
Advisors: Dennis K. Pearl, Matthew Beckman, Priyangi Bulathsinhala, Robert Carey, Kathleen Smith

2017 Undergrads: Alex Chen, Qichao Chen, Jinglin Feng, Zibin Gao, Sitong Liu, Ryan Manigly-Haney, David Robinson, Yingjie Wang, Caihui Xiao, Yuxin Zhang

2018 Undergrads: Jiajun Gao, Stephen Li, Thomas McIntyre, Samuel Messer, Angela Ting, Ryan J Voyack, Luxin Wang, Zhiliang Zhang, Yinqi Zhang, Yubaihe Zhou

Book of Shiny Apps for Statistics Teaching

Variable Matching Game



Variable Types Home Level 1

Drag the variables into the categories they belong to. you have used: 155 secs

Family Size (1,2,3,...) House's number of bedrooms Occupation Stock quotes for a stock over days Result of rolling a die Blood Pressure

Type of schools (public, private) Grade in course (A,B,C,D,F) Age Air Pollution Index (0-50,...,300+) Marital Status Width (inches)

Stage of disease Sleep time to nearest hour Social economic class Brand of Shoe

Quantitative & Discrete: [] [] [] []

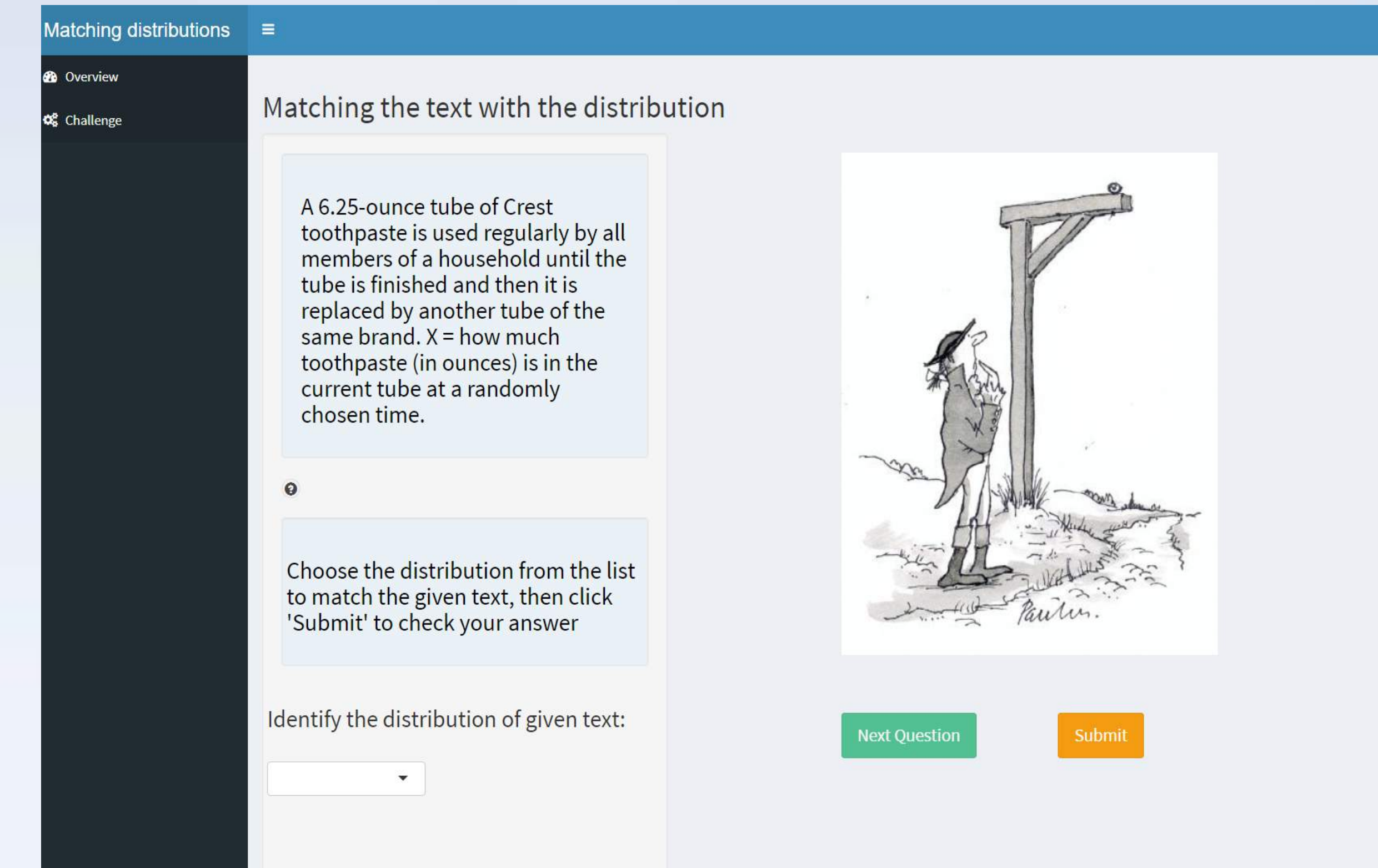
Quantitative & Continuous: [] [] [] []

Qualitative & Nominal: [] [] [] []

Qualitative & Ordinal: [] [] [] []

[[Previous]] [Next]]

Matching Distribution




Matching distributions

Overview Challenge

Matching the text with the distribution

A 6.25-ounce tube of Crest toothpaste is used regularly by all members of a household until the tube is finished and then it is replaced by another tube of the same brand. X = how much toothpaste (in ounces) is in the current tube at a randomly chosen time.



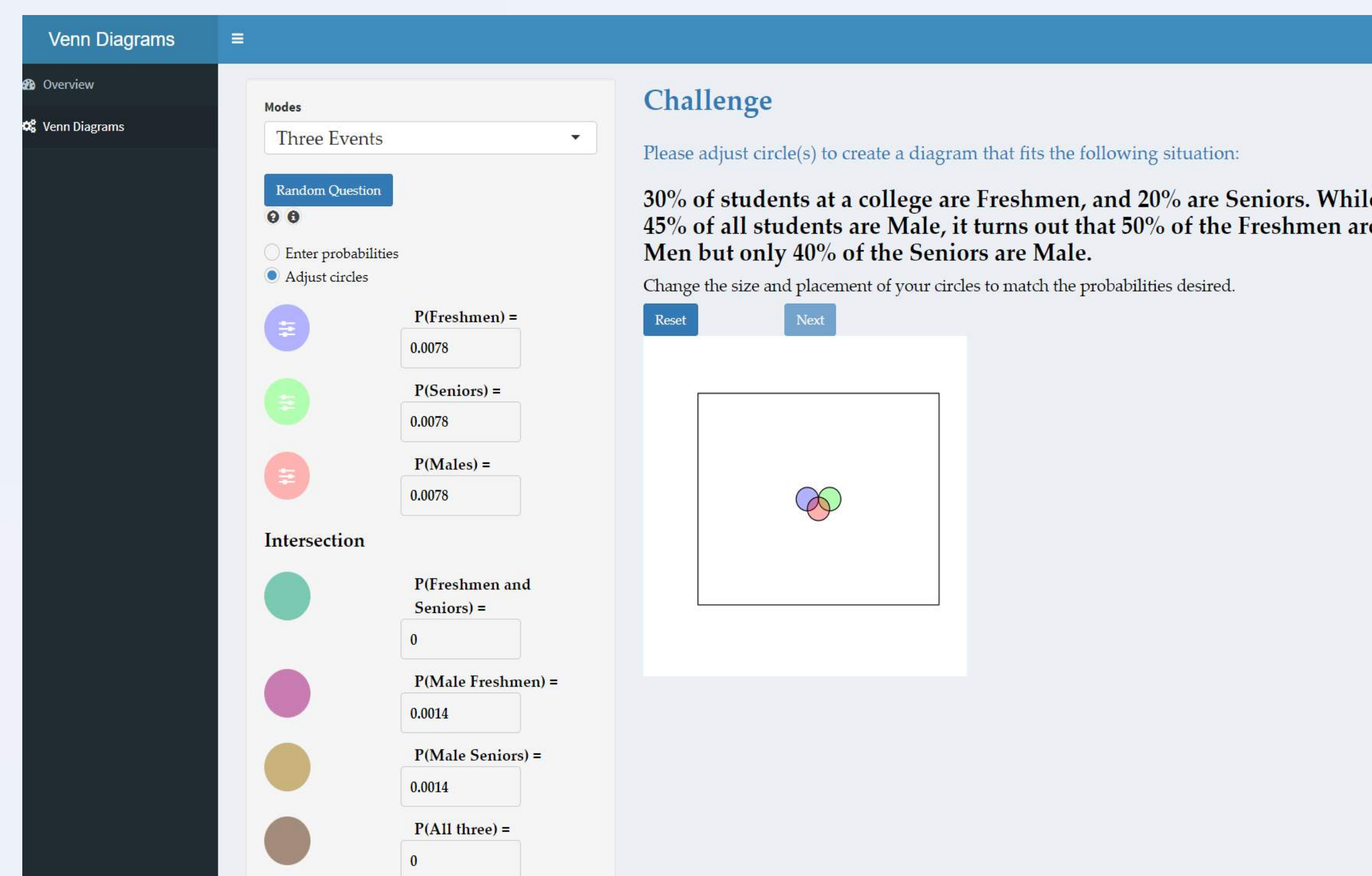
Choose the distribution from the list to match the given text, then click 'Submit' to check your answer

Identify the distribution of given text:

[]

Next Question Submit

Venn Diagrams Game



Venn Diagrams

Overview Venn Diagrams

Modes: Three Events

Random Questions

Enter probabilities Adjust circles

Challenge

Please adjust circle(s) to create a diagram that fits the following situation:

30% of students at a college are Freshmen, and 20% are Seniors. While 45% of all students are Male, it turns out that 50% of the Freshmen are Men but only 40% of the Seniors are Male.

Change the size and placement of your circles to match the probabilities desired.

Reset Next

$P(\text{Freshmen}) = 0.0078$
 $P(\text{Seniors}) = 0.0078$
 $P(\text{Males}) = 0.0078$
 $P(\text{Freshmen and Seniors}) = 0$
 $P(\text{Male Freshmen}) = 0.0014$
 $P(\text{Male Seniors}) = 0.0014$
 $P(\text{All three}) = 0$

ANCOVA



ANCOVA

Select the time limit (second): 60

Time left: 1M

Start

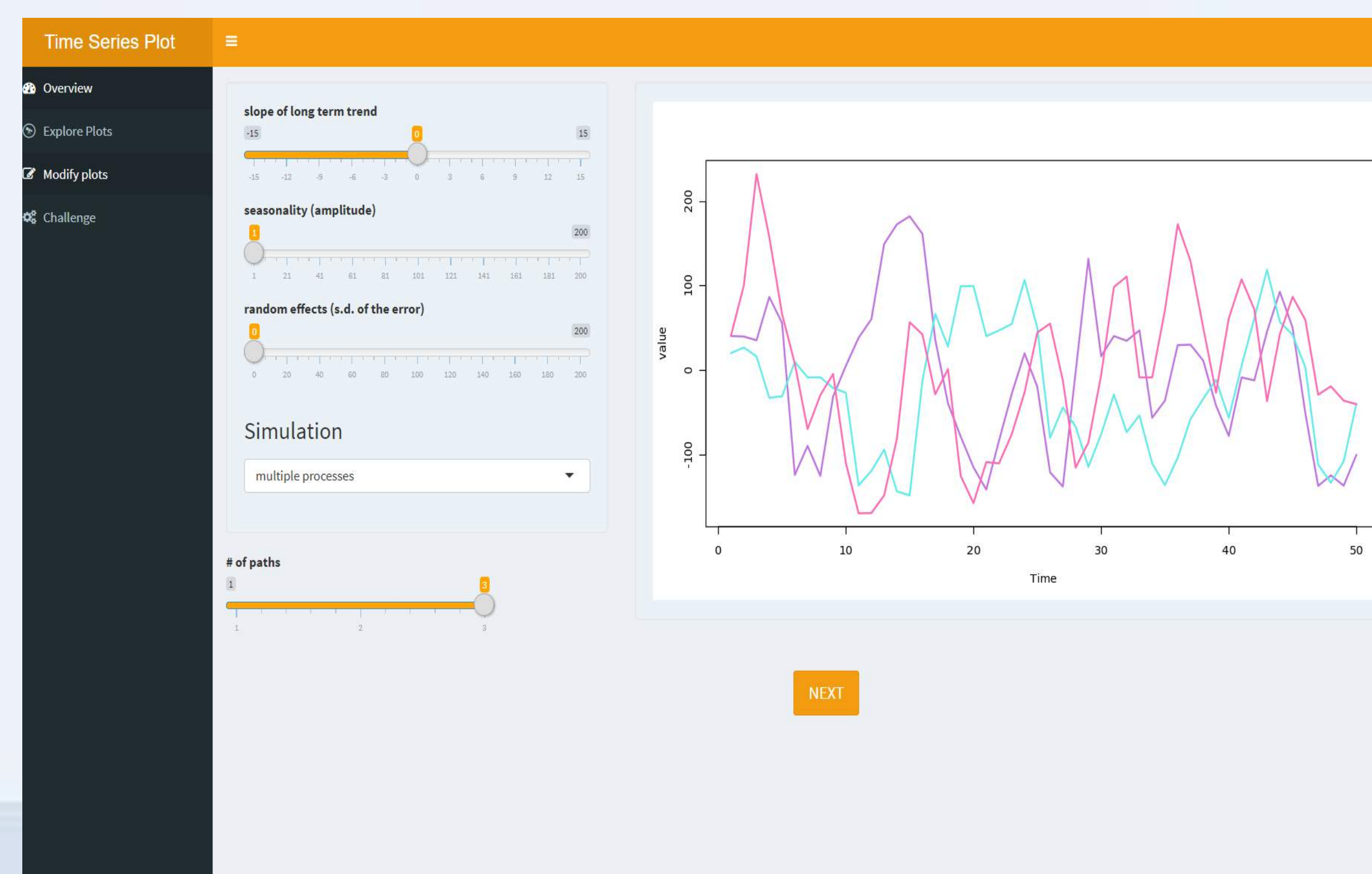
Three scatter plots (A, B, C) showing data points and regression lines.

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X	1	107916	107916	1191.618	<2e-16 ***
X:cov	1	1321637	321637	3580.822	<2e-16 ***
Residuals	296	35570	90	1.143	0.2857

Submit Answer

Time Series Plot Game



Time Series Plot

Overview Explore Plots Modify plots Challenge

slope of long term trend

seasonality (amplitude)

random effects (s.d. of the error)

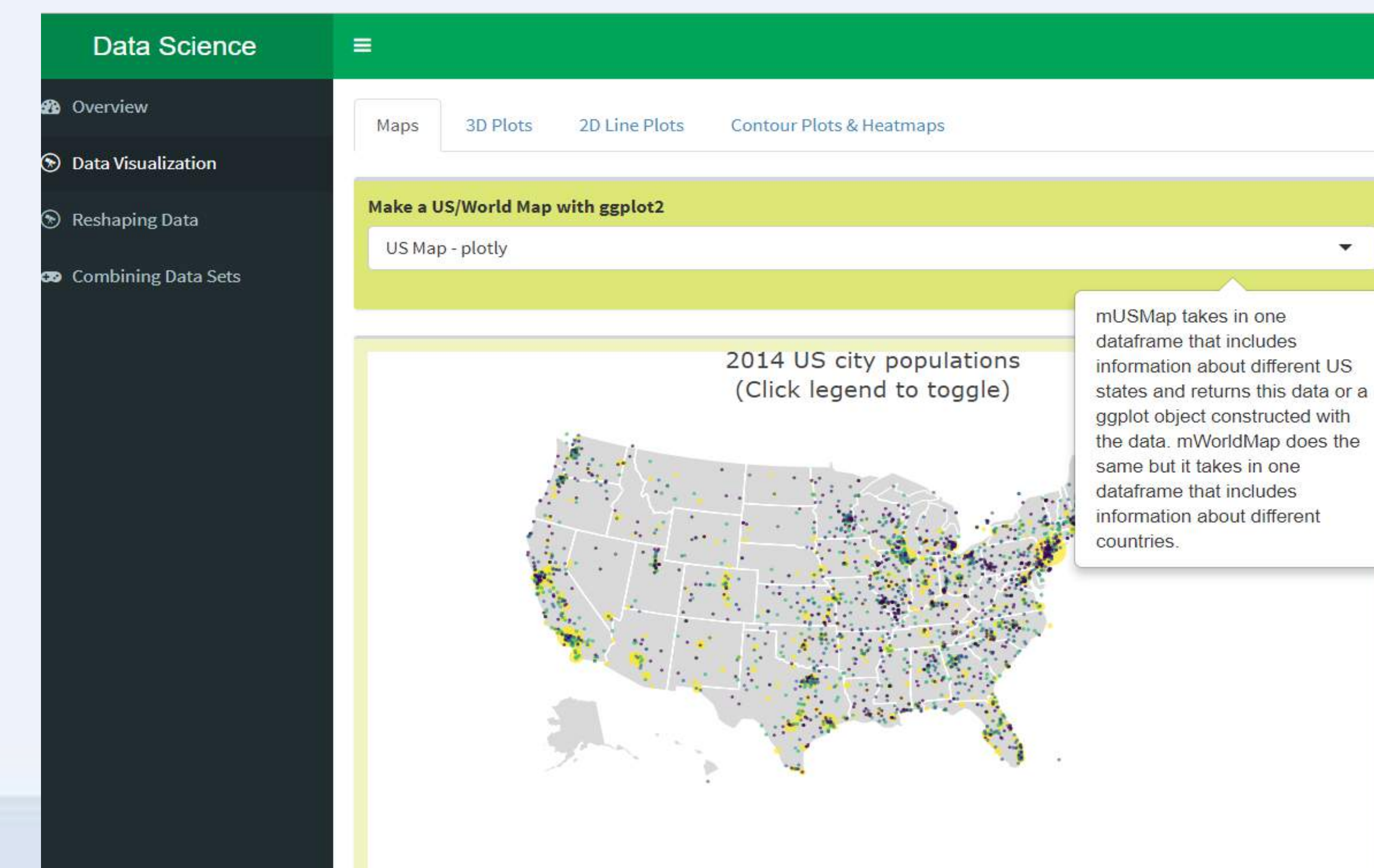
Simulation: multiple processes

of paths

Time Series Plot showing multiple paths over time.

Next

Data Science



Data Science

Overview Data Visualization Reshaping Data Combining Data Sets

Maps 3D Plots 2D Line Plots Contour Plots & Heatmaps

Make a US/World Map with ggplot2

US Map - plotly

2014 US city populations (Click legend to toggle)

mUSMap takes in one dataframe that includes information about different US states and returns this data or a ggplot object constructed with the data. mWorldMap does the same but it takes in one dataframe that includes information about different countries.