StatPREP Newsletter

OCTOBER 2021 | ISSUE #24

A FRESH APPROACH TO INTRO STATISTICS

by Donna LaLonde

"The title of this article is how the authors of a new book describe their approach. Introduction to Modern Statistics (IMS) is a new book authored by Mine Çetinkaya-Rundel and Johanna (Jo) Hardin. Key components of this book are its early introduction of multi-dimensional thinking and the use of both simulation techniques and traditional methods. The text is supported by supplementary materials including videos, tutorials, labs, and datasets. The book is a part of the OpenIntro (https://www.openintro.org/) which is committed to lowering the barriers to education by making books freely available online.

IMS has 27 chapters organized into six parts – Introduction to data, Exploratory data analysis, Regression modeling, Foundations of inference, Statistical inference, and Inferential Modeling and has two appendices – Exercise solutions and References. Each section concludes with an application. The application presents a case study, tutorials built using learnr and supporting self-paced study, and a lab which uses R to explore the data analysis described in the case study. The authors describe this book as a "re-imagining" of Introduction to Statistics with Randomization and Simulation with a "heavy emphasis" on exploratory data analysis.

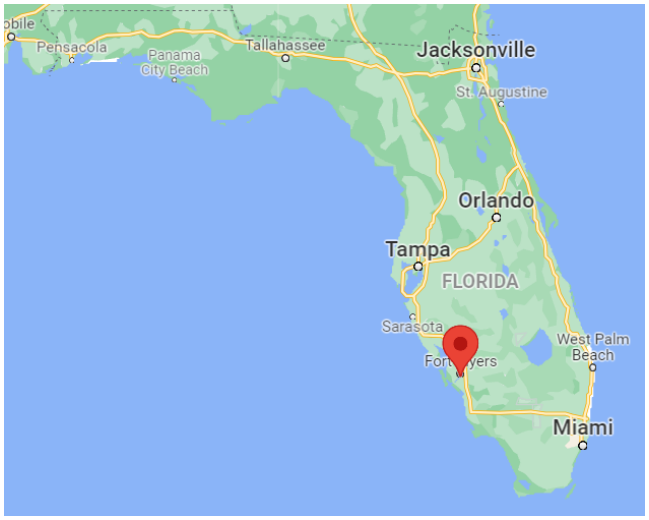
In the preface the authors share their goals for the readers of this text. They want readers to appreciate the wide-range of practical applications, to appreciate that everyone can learn from data – "you don't have to be a math guru," and that data are messy and statistical tools are imperfect, so it is important to understand the strengths and weaknesses of the tools.

Please check out this resource and share your insights on the StatPrep community!

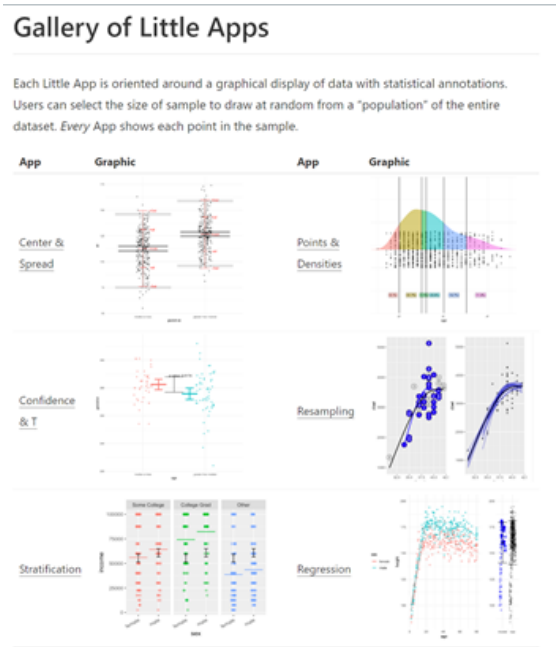
COMING SOON: STATPREP SUMMER 2022 WORKSHOP APPLICATIONS

by Ambika Silva

After what feels like a very long time, StatPREP is pleased to promote their return to an in-person workshop. We will be holding an in-person workshop in Fort Myers, Florida, located in the southwest region of Florida.



Statistics courses should be data centric! This workshop will expose participants to free online tools from StatPREP (NSF DUE-1626337) that can help your students learn fundamental concepts in introductory statistics. Hands-on computer activities will guide participants in use of some of the Little Apps activities featured at http://www.statprep.org/LittleAppSite/.



Participants will also get the opportunity to explore tutorials to introduce RStudio then actually get started using RStudio on their own! Whether you're new to RStudio and programming or have been using it for years, we believe you will benefit from this exciting workshop session.

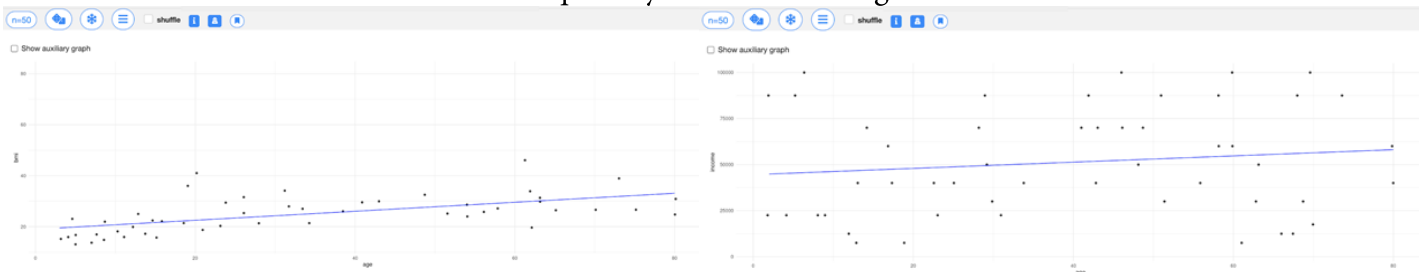
This is a great workshop to attend to learn how to incorporate real data into your course. In a world where we may be teaching both online and in person, this workshop provides ways to satisfy all modalities of education using free technology. The application portal will be open November 2021 and will be found at StatPREP.

LITTLE APP ACTIVITY: INTRODUCING LINEAR REGRESSION

By Kathryn Kozak

Regression is all about visualizing the relationship between two (or more) variables. So it really helps to have students be able to look at scatter plots of different variables, so they can see when two are related linearly or with a different relationship. It is also important for them to see how strong the relationship is. The Little App Linear Regression is very useful for both of these visualizations. The activity Introducing Linear Regression is very helpful in leading students to explore some variables.

The activity starts by giving students examples of variables to look at, and describes how strong the relationships are. The first variables to explore are BMI and age. The activity gives links to BMI calculators so students can learn what this quantity is. This allows a great life lesson for the students.



As you can see from the graph, there is a fairly strong positive linear relationship. This gives students a visual representation of a strong relationship between variables. Next though the students look at BMI versus income. This graph would look something like this.

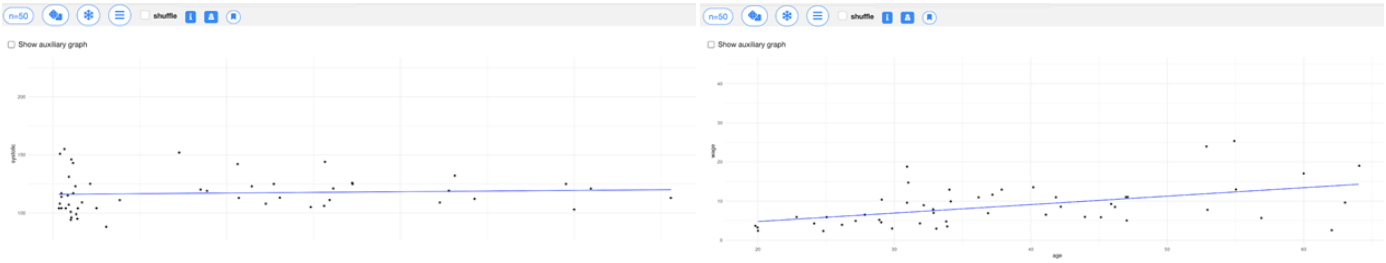
In this case, there doesn't seem to be a linear relationship. The activity tells the students this so they can learn what a linear relationship looks like and what one does. As the students work through other variables to use as the explanatory variable along with BMI, they are exposed to negative relationships, outliers, and relationships of varying strength.

Once the students have been exposed to the types of relationships, then they switch to different datasets and variables. Here the students aren't told what kind of relationship there is and they can start to use their knowledge to determine on their own.

After this exploration, the students are asked to find their own explanatory variable to explain systolic blood pressure. So they need to look through the variables and think of what explanatory variables affect the response variable of blood pressure.

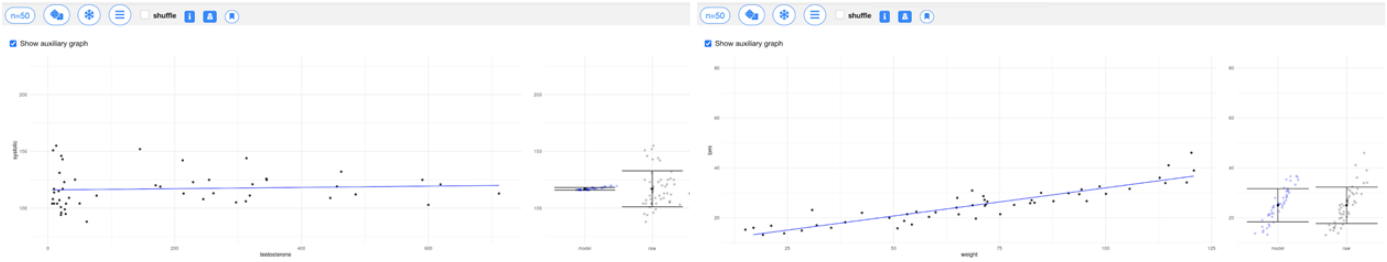
It is interesting for the teacher to see what ideas students come up with. If the timing works, you could ask the groups to share the variables they identified as explanatory.

....continued on page 4



After this exploration, students move on to finding a way to visualize the coefficient of determination, or R-squared. R-squared is a better way to look at how much of the response variable is explained by the explanatory variable. Students can really see how one variable explains more than others. The activity steps the student through this visualization. The advantage of R-squared is that it can work if you have a covariate variable, for which correlation is only applicable for one response and one explanatory variable. If you have time, you can expand this activity by having students explore on their own by picking their own variables to compare.

Here are a couple of examples of regression graphs with the variability of the model and the raw data displayed. Students can see how R-squared is the ratio of how much variability in the response variable is explained by the model. The first graph has a fairly low R-square. The second has a much higher one. This should help your students understand the strength between two variables.



Again, if you have time, you could ask the groups to share what they found.

Along with R-squared, the activity also has the students explore effect size, which for regression is the slope. Again the activity has the students estimate the effect size and the R-squared value for two different response variables with explanatory variables that interest them. This really has the students own their learning since they can find variables that they can relate to. Lastly, the students explore high prices using a different dataset.

This activity is a very useful way to introduce linear regression. You can expand this activity by having your students pick their own response and explanatory variable. I hope you find this activity will help your students understand the concept of regression.

Upcoming Webinars

INTRODUCTION TO RSTUDIO AND EXPLORATORY DATA ANALYSIS

Presented by Joe Roith & Kathryn Kozak

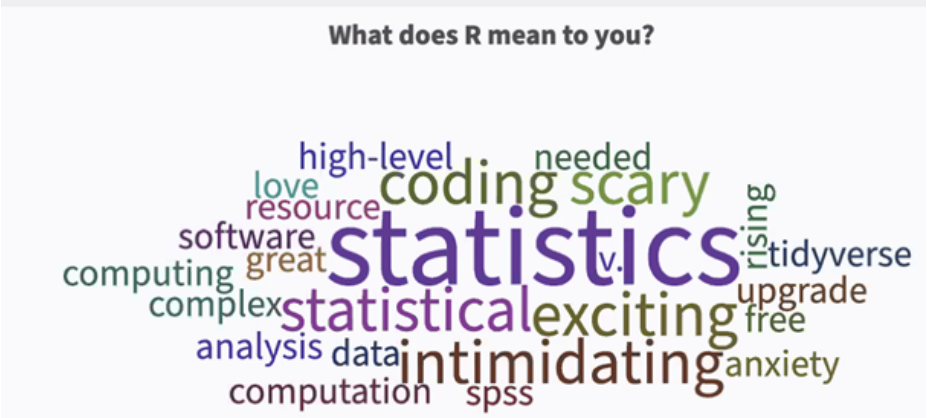
Monday, November 8
4:00pm ET

This workshop will introduce participants to RStudio using RStudio Cloud. In addition RStudio server and RStudio desktop will be described. This introduction will include packages and reading data into RStudio. Then there will be an introduction to Exploratory Data Analysis (EDA).

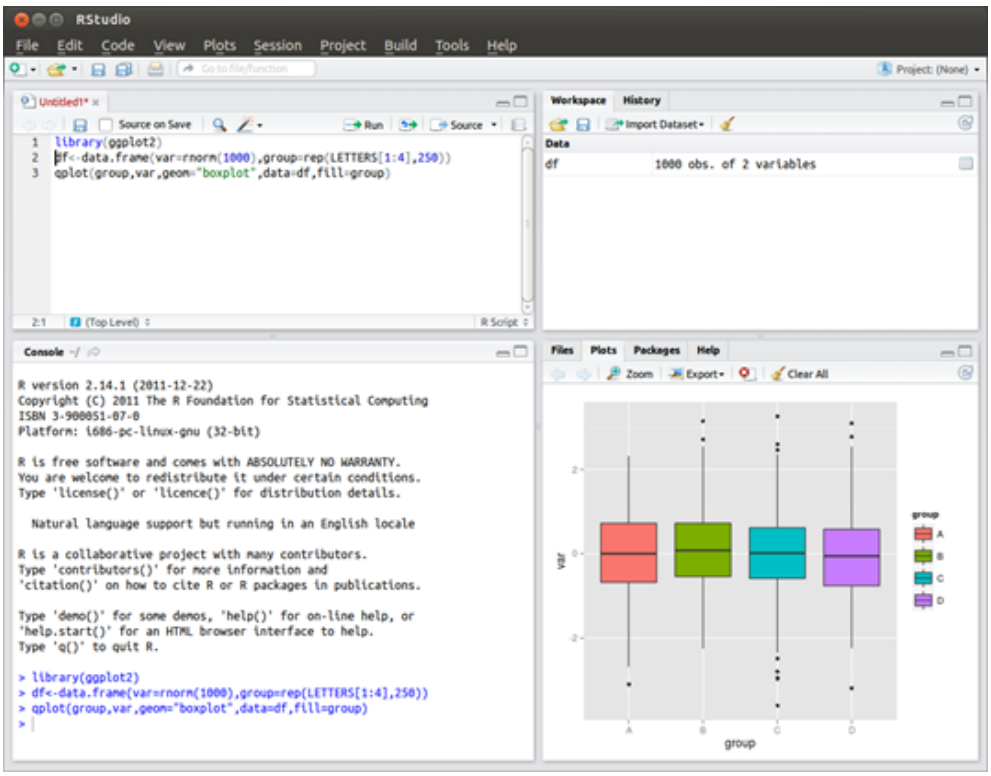
REGISTER

Want to Learn about RStudio?

At a recent workshop, we asked participants "What word comes to mind when the software R is mentioned?" The results are below and contain a wide variety of thoughts and emotions.



RStudio is an IDE, or integrated development environment, for R programming. Download and install it from http://www.rstudio.com/download . RStudio is updated a couple of times a year and should let you know when you need to update. RStudio has beautiful graphics, powerful tools, and is free to use!



Why choose RStudio to use? Some would argue it allows for a better workflow, that it's easier on reproducibility. As an instructor I like that it's free for students to use and that it's a cloud or server based product. If you're interested in an introduction to RStudio I encourage you to attend the upcoming webinar presented by Joe Roith and Kate Kozak.

While videos and webinars are great for some, others learn best by curling up with a good book. R For Data Science is available for free as an online book at https://r4ds.had.co.nz. Or better yet… do both!!!!

Save the Date

AMATYC CONFERENCE 2021

Session: S091, Title: Utilize StatPREP's Free Little Apps to Teach Data-Centric Statistics
Day: Saturday, October 30
Time: 10:45 am - 11:35 am
Room: Deer Valley
Presenters: Helen Burn, Carol Howald, Rona Axelrod & Joe Roith

Learn about free, online tools from StatPREP that can help you and your students learn fundamental concepts in introductory statistics. StatPREP is an NSF-supported faculty development project sponsored by the Mathematical Association of America (NSF DUE-1626337). Faculty leaders from five national StatPREP hubs will share their use of the menu-and-slider Applets made with R markup language.

LEARN MORE

STATPREP SUMMER 2022 WORKSHOP

In-Person Workshop in Fort Myers, Florida, will be held on May 20-21, 2022.

Details on how to register will be coming!