

October  
30, 2019

# StatPREP



# MAA

MATHEMATICAL ASSOCIATION OF AMERICA

## NEWSLETTER

---

### A REVIEW OF A REVIEW

BY DANIEL KAPLAN

---

This essay started out as a review of David Spiegelhalter's estimable new book, "The Art of Statistics: Learning from Data." Then it turned into a review of a New Yorker book review by Hannah Fry. (See <https://www.newyorker.com/magazine/2019/09/09>.) Fry, whose own book "Hello World: Being Human in the Age of Algorithms" is on the StatPREP short list for book reviews, had a captivating lede for her review of "The Art of Statistics." Much of the review is about Bayesian statistics and the current unravelling of the p-value as a scientific tool.

In her review, Fry wrote briefly about "a Belgian astronomer and mathematician named Adolphe Quetelet." Quetelet plays an outsized role in the history of statistics. He coined the term "average man" ("l'homme moyen) and pioneered the application of means and rates to social phenomena. This seems blindingly obvious now, but in the 1830s crime and such were seen as unrelated individual events, rather than the outplay of social forces. Quetelet also invented the "Quetelet Index," which is now known as the body mass index (BMI). I wondered why Fry didn't follow the usual practice of describing polymath Quetelet as an astronomer, STATISTICIAN, mathematician, and SOCIOLOGIST. Was this an oversight? That would be a surprise given how well Fry has written about statistics. And Fry's description turns out to be historically correct when it comes down to Quetelet's invention of social statistics.

Quetelet was indeed an astronomer, not a statistician (yet). From Brussels, he was sent to Paris to study the organization of the prestigious Paris observatory with an eye to setting up such an institution in Belgium. The late 1820s, when Quetelet's arrived in Paris, was the end of a long period of French national reorganization after the chaos of Napoleon's resurrection and collapse at the Battle of Waterloo. In particular, France was being organized into the state-like départements which exist to this day. A mundane part of this re-organization was the harmonization of government record-keeping across the départements. Due to this, extensive records covering the whole nation were being collected and published in Paris.

*Continued on page 2*

#### WHO'S WHO:

##### LEADERSHIP TEAM

Mike Brilleslyper,  
Air Force Academy

Jenna Carpenter,  
Campbell University

Danny Kaplan,  
Macalester College

Kathryn Kozak  
Coconino Community  
College

Donna LaLonde,  
ASA

Ambika Silva  
College of the Canyons

Rachel Levy  
MAA

##### HUB LEADERS

Joe Roith, St. Olaf's Col-  
lege, Northfield, MN (2017-  
18)

Ambika Silva, College of the  
Canyons, Santa Clarita, CA  
(2017-18)

Helen Burn, Highline Col-  
lege, Seattle, WA (2018-19)

Hwayeon Ryu, Elon Univer-  
sity, Elon, NC (2018-19)

Carol Howald, Howard Com-  
munity College, Columbia,  
MD (2019-2020)

Thomas Kinzeler, Tarrant  
County College, Fort Worth,  
TX (2019-2010)

Rona Axelrod, Florida SW  
State College, Fort Myers,  
FL (2020-2021)

Brooke Orosz, Essex Coun-  
ty College, Newark, NJ  
(2020-2021)

Support for this MAA Program is provided by NSF DUE-1626337



---

## A REVIEW OF A REVIEW continued...

---

These records were a new kind of thing; it wasn't clear what purpose they solved other than being a small part of the transition in France to a centrally governed country. Quetelet found a use for them, applying astronomical techniques to summarizing the sprawling mass of numbers. For "astronomical technique," read "the mean," but before Quetelet it was a method used almost entirely in astronomy and geodesy. Quetelet found that mean crime rates (among other social phenomena) were surprisingly constant from one year to the next, but also differed consistently from one département to another. He concluded that the level of crime is set by the social characteristics of each département.

*"Data precedes analysis" -  
Daniel Kaplan*

There's an important lesson here for us today that goes beyond the intrinsic interest in history. Put in contemporary terms, the French state in the 1820s constructed the first consistent data warehouse of social events. Companies, government, and other institutions often create such warehouses now, simply because computers make it easy to do and because it's widely believed that even incidental data might reveal something about how things work. We hear many stories today about successful and useful discoveries based on incidentally collected data. (We don't hear about the failures, but that's because nobody wants to publicize their failures.)

The lesson? Data precedes analysis. Statistical methods, such as Quetelet's introduction of the mean, are motivated by the needs of summarizing and modeling data. A more famous example of this occurred a half century after Quetelet's invention. Francis Galton had data on the heritability of traits in humans and sweet peas, but had to work to find a way to make sense of it, inventing a string of methods leading in the end to his "co-relation coefficient." It's not until Fisher, a mathematician forced to deal with field research, that we see an example of method preceding data: Fisher's invention of experimental methods such as Latin Squares and ANOVA that guide how to collect data.

So, as you teach statistics, keep in mind that data is not in service to theory. It's the other way round: the statistical pioneers invented theory to respond to the demands of data. Wouldn't it make sense to lead students from data to methods, rather than teaching methods as an abstraction and data merely as the numbers to plug into the methods.

## SAVE THE DATE!

### Summer 2020 StatPREP Workshops

Fort Myers, FL: May 29 - 30, 2020, Florida Southwestern State College

Fort Worth, TX: June 3 - 4, 2020, Tarrant County College

New Jersey: June 5 - 6, 2020, Essex County College

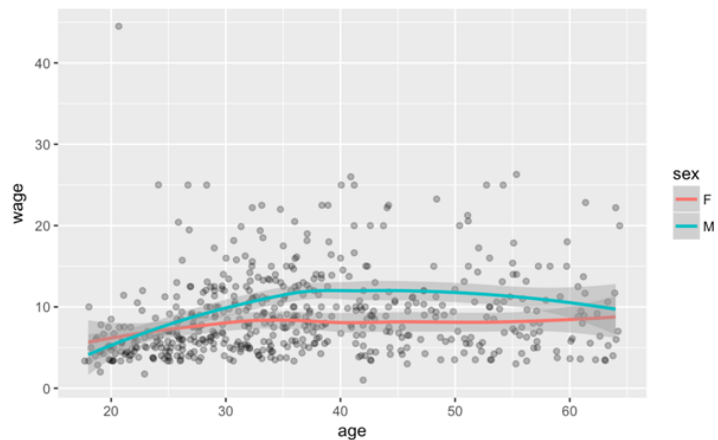
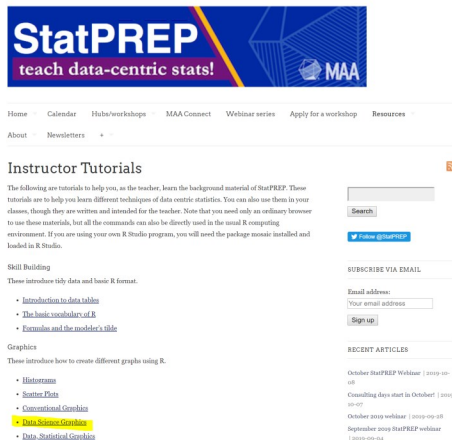
Columbia, MD: June 12-13, 2020, Howard Community College



# DISCOVERING StatPREP.org

BY AMBIKA SILVA

Have you checked out the “Instructor Tutorials” page? You can find this under the “Resources” tab on the main page of the StatPREP website, or by going to <http://statprep.org/resources/>. These tutorials can help instructors learn different techniques to teach data-centric statistics. Several exciting graphics can be downloaded to use in your classroom as a discussion topic or be the focus of a worksheet or other assessment. For example, the following graphic came from the Resources tab > Instructor Tutorials > Graphics > Data Science Graphics (note wage is in \$/hour).



**Questions you could ask your students about this graphic include:**

**1. At what approximate age(s) are males and females making about the same wage?**

Answer: Approximately 23 years old

**2. At what age(s) is there the most substantial gap in wage difference between males and females?**

Answer: Approximately 38 -45 years old there seems to be the largest gap

**3. What age ranges does it appear that females make more than males and vice versa?**

Answer: Females make more when they are approximately 16 - 23 years old, men make more between 23 and 64 years old

**4. Is it possible that Amy, who is female, makes more than Greg, who is male if they are both 40 years old?**

Answer: Yes, it's possible!

*I encourage you to explore and find a graphic that you think would be a good discussion for your own class, then post it on MAA Connect (<https://connect.maa.org/home>)! - Ambika Silva*

**Want more StatPREP? Check out:**

**<http://statprep.org/>**

---

## **A REVIEW OF "ASK GOOD QUESTIONS"**

### **BLOG BY ALLAN ROSSMAN**

**BY KATE KOZAK**

---

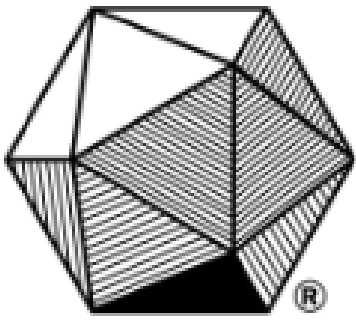
You may have encountered some of the excellent statistics textbooks Allan Rossman has co-authored: "Workshop Statistics," "Statistical Investigations" (which, along with the Lock5 text is a leading text introducing statistics through resampling, randomization, and bootstrapping), and "Stat 2: Building Models for a World of Data." Allan is also famous as the master of ceremonies of the US Conference on Teaching Statistics (USCOTS), which he leads with his distinctive west-Pennsylvania lilt.

"Ask good questions," is Allan's teaching motto. He's now launched a web blog at <https://askgoodquestions.blog/> to help all of us benefit from his experience asking good questions of his students at Cal Poly San Luis Obispo, one of the leading undergraduate statistics programs in the country. Every week, Rossman posts questions that he asks his statistics class to give you ideas of questions you can ask your class.

In a recent post, Allan posed a question for instructors: How confident are you that your students can interpret a 95% confidence interval (CI) correctly? This post notes that it is a continuation of a previous one considering numerical data and highlighting a common misconception about interpreting a CI for a population mean.

You'll also find activities on the blog that you can bring to your class to help you infuse data-centric ideas into your statistics lessons. The 16 essays already posted are a valuable resource. Read them to get some great ideas for your class. And be sure to check back every week for the latest posts!





# MAA CONNECT

Do you have questions but no one to ask? Have you tried using a Little App and you want to share your experiences with others? Do you want advice on how to adjust your curriculum to be more data centric? There is a place to have all of these discussions. It is called MAA Connect. The MAA Connect website is <https://connect.maa.org/home>. You do not need to be an MAA member to be part of this community.

Once you log into MAA Connect, go into My Communities. At that point, you can select the StatPREP Hub Community. In this community you can post a discussion, send private messages to another member, and post materials to the library. There are already discussions you can join, so feel free to look through what is already there. There are also materials posted in the library, including:

- Assessment ideas for Introductory Statistics Classes
  - Getting started with RStudio Cloud
  - Links to Videos using R Studio
  - PeerJ Data Science Collection
    - Sample Syllabus
- What your doctor should know about statistics

Once you are a member of this community, you can see all of this material. Contact Kathryn Kozak at [kathryn.kozak@coconino.edu](mailto:kathryn.kozak@coconino.edu) for any questions, assistance, or to join the MAA Connect StatPREP Hub Community.

---

# TEXTBOOK REVIEW:

## PROJECT MOSAIC LITTLE BOOKS

BY DONNA LALONDE

---

Quality and cost are two important criteria for choosing textbooks and often quality comes at a significant cost. The Project Mosaic Little Books are an exception. They are high quality resources at no cost. The books are freely available both for online viewing and download. The Little Books - Start Teaching Statistics with R, A Student's Guide to R, and Start R in Calculus are available at <https://github.com/ProjectMOSAIC/LittleBooks>.

On the website Project Mosaic is described as "a community of educators working to develop a new way to introduce mathematics, statistics, computation and modeling to students in colleges and universities." The effort originated in 2009 with funding from the National Science Foundation. Although the grant has ended, the community remains active. To learn more about Project Mosaic visit <http://mosaic-web.org/>.

The authors of Start Teaching with R will be familiar to StatPREP workshop participants as they have also been presenters. In this "little book," Randall Pruim, Nicholas Horton, and Daniel Kaplan share their approach to teaching statistics with R. The approach the authors have taken with the book make it an excellent resource or complement to be used with an existing text. The book is designed both to provide classroom-ready projects and to help instructors gain new expertise. The teaching

tips that are included as margin notes are succinct and will be useful for lesson planning. Since the materials are copyrighted under a Creative Commons 3.0 Attribution License, instructors may share and use the materials to develop their own projects as long as attribution to the original work is provided.

A Student's Guide to R also written by Horton, Pruim, and Kaplan was designed to complement Start Teaching with R. It is a reference book covering topics like - getting started with R Studio, quantitative and categorical variables, and data wrangling. Since this book was designed as a reference book, the margin notes guide the student to sources of additional information on a given topic. A Spanish translation of A Student's Guide to R is available at <https://github.com/jarchoeltrocho/MOSAIC-LittleBooks-Spanish>.



# UPCOMING EVENTS

## NOVEMBER 2019

SUN	MON	TUE	WED	THU	FRI	SAT
					1	2
3	4	5 Webinar	6	7	8	9
10	11	12 Consulting Session	13	14	15 Consulting Session	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

## DECEMBER 2019

SUN	MON	TUE	WED	THU	FRI	SAT
1	2	3	4	5	6	7
8	9	10 Consulting Session	11	12	13	14
15	16	17	18	19	20 Consulting Session	21
22	23	24	25	26	27	28
29	30	31				



**MAA**  
MATHEMATICAL ASSOCIATION OF AMERICA



**ASA**