Civic Data Hacking Using R and RStudio

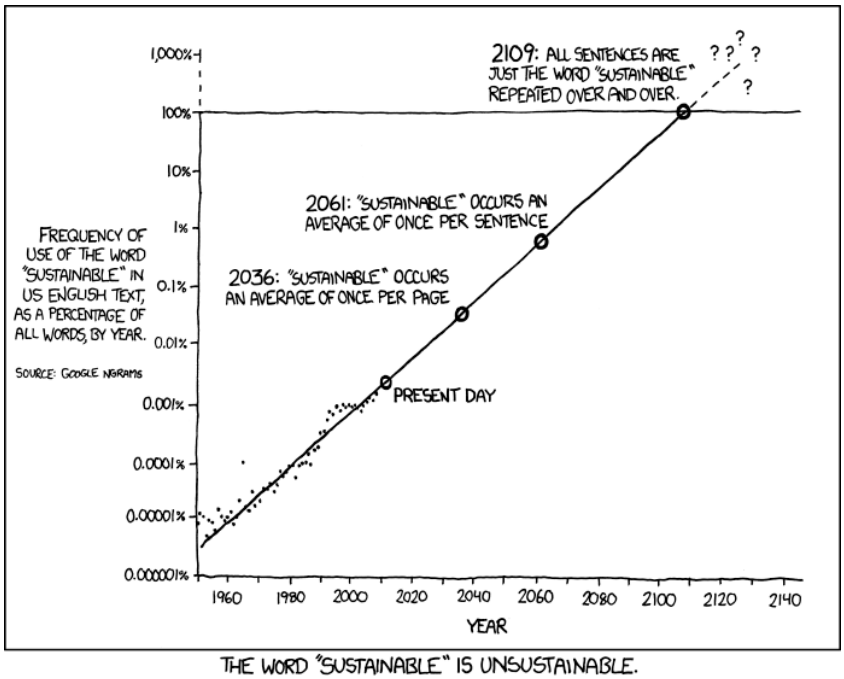
Joy Payton
https://www.linkedin.com/in/joypayton/

"Data can be the source of data journalism, or it can be the tool with which the story is told — or it can be both. Like any source, it should be treated with scepticism; and like any tool, we should be conscious of how it can shape and restrict the stories that are created with it."

— Paul Bradshaw, Birmingham City University
http://datajournalismhandbook.org/1.0/en/introduction_0.html

"Citizen science bridges gaps by harnessing the power of people who are motivated by curiosity, a desire to advance research, or a concern about environmental conditions in their communities, then connecting them to projects that benefit from their energy and dedication."

https://scistarter.com/citizenscience.html



THE WORD "SUSTAINABLE" IS UNSUSTAINABLE.

https://www.xkcd.com/1007/

"The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data."

— John W. Tukey , "Sunset Salvo", *The American Statistician* 40(1), 72-76 (February 1986)

My role today is to provide you with tools and a simple example. You will not be a data scientist by the end of my presentation, but you may have garnered enough bare-bones jargon to Google things effectively, and that, combined with your increased confidence that you really can do this, along with the burning curiosity or disquiet you have about [***that thing you're so interested in***] are enough to make you a competent citizen scientist. Welcome!

To try R without downloading anything:

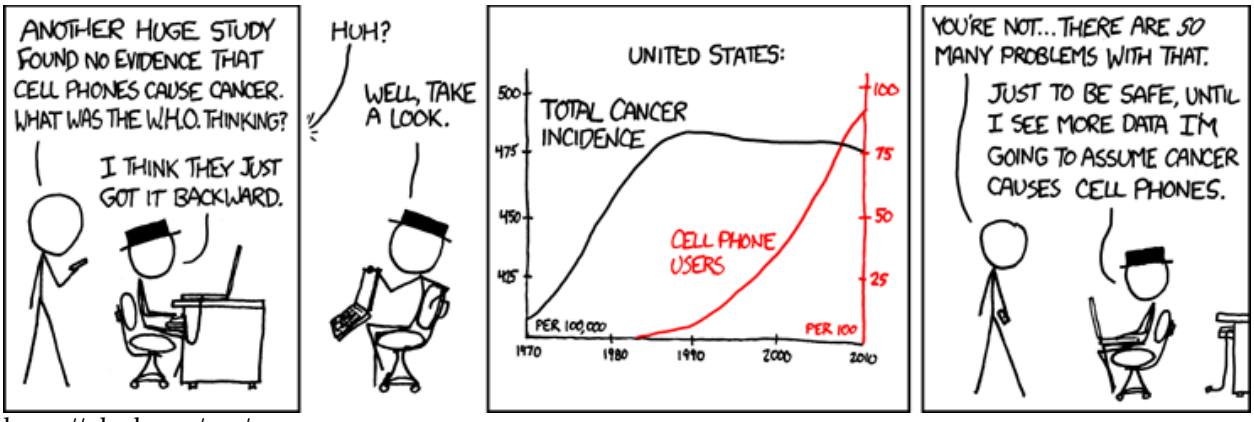
It's important to start with RStudio as soon as possible. But if you want to try R briefly without downloading R or RStudio, I suggest the free Intro to R course at DataCamp (https://www.datacamp.com/courses/free-introduction-to-r). All of the DataCamp courses I've seen are great, but you don't get the RStudio experience.

Alternatively, you can also try RStudio Cloud (which at the time of this writing is still in alpha, so no promises) at https://rstudio.cloud/. Again, you get a chance to try it out without downloading anything.

To get started with R on your workstation:

Download R (https://www.r-project.org/) and RStudio Desktop (https://www.rstudio.com/products/rstudio/download/) . You'll also want to install a few central packages using install.packages("name"). I suggest starting with:

- tidyverse (it's a biggie)
- knitr
- swirl
- leaflet
- ggthemes
- ggThemeAssist
- ggmap
- effsize
- scales
- printr
- kableExtra



https://xkcd.com/925/

To learn basic inferential statistics:

This is important if you really want to become a citizen scientist. One possibility is using a Swirl course (https://swirlstats.com/) to learn stats within R itself. This may not be enough scaffolding for someone new to science and statistics, however. If you like more structure, I highly recommend Khan Academy (https://www.khanacademy.org/math/statistics-probability) for learning statistics. You should at least learn:

- Probability distributions
- Central limit theorem
- T tests / Hypothesis testing
- Confidence Intervals
- ANOVA
- Chi-Square
- Linear regression

One more thing to consider:

Check out RPubs (https://rpubs.com), which allows you to make your discoveries public easily.