

Introduction To Data Science!

Course Website!

We have a course website here:

<https://grantinnerst.github.io/MAT-219>

This will hold all of the course information (slides, readings, hw, projects, etc.)

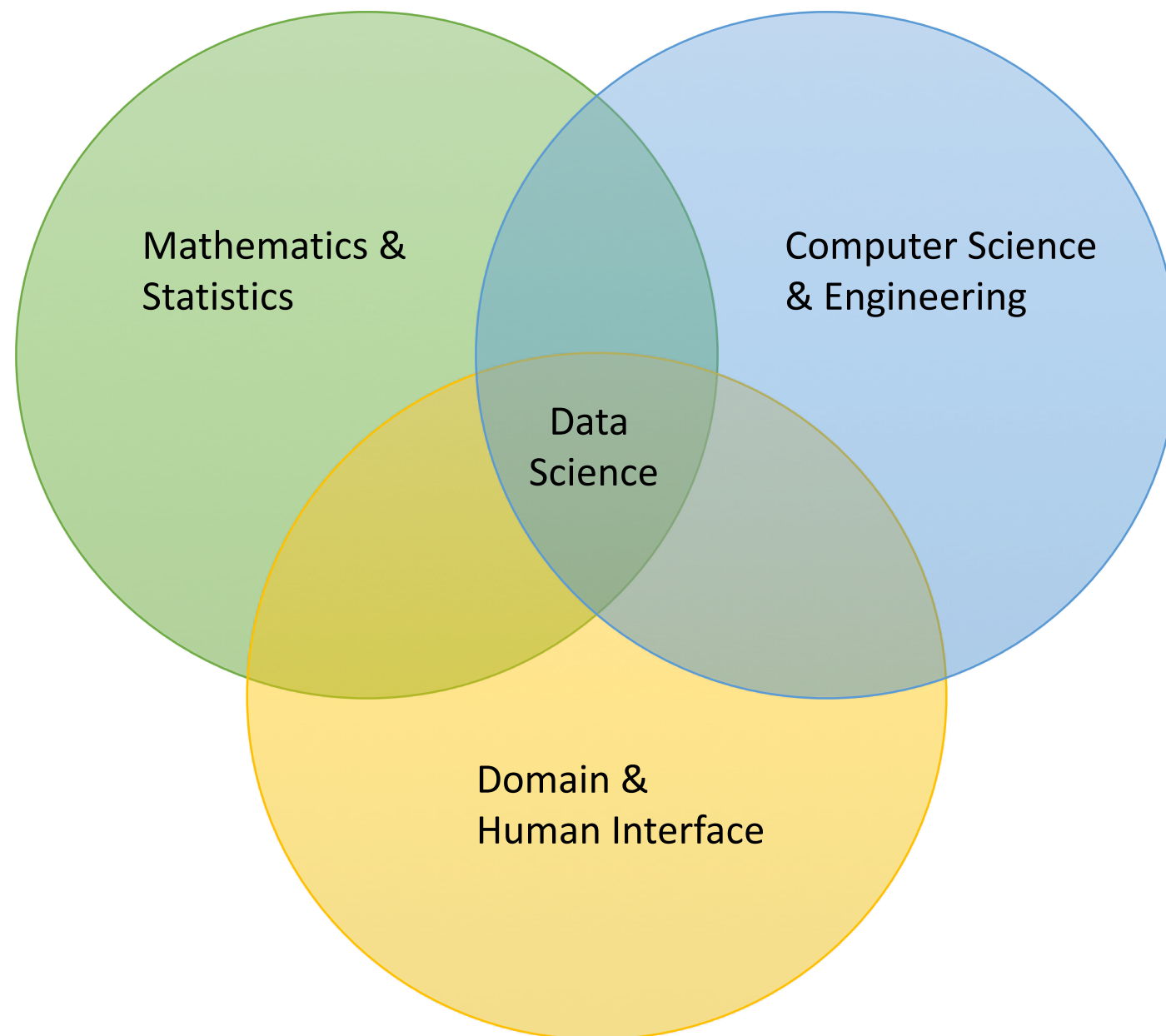
It also has helpful links to our book(s) and other tools we will use daily!

I will update this site as much as possible!

What is Data Science?

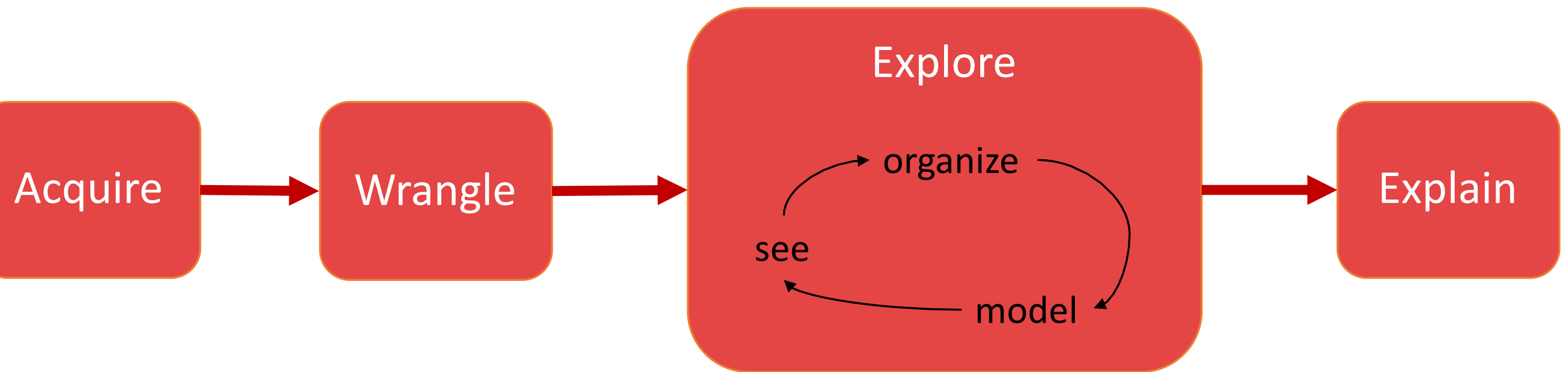
Data science is an exciting discipline that allows you to turn raw data into understanding, insight, and knowledge!

We are going to do data science in a *tidy* way with an emphasis on statistical thinking.



What is Data Science?

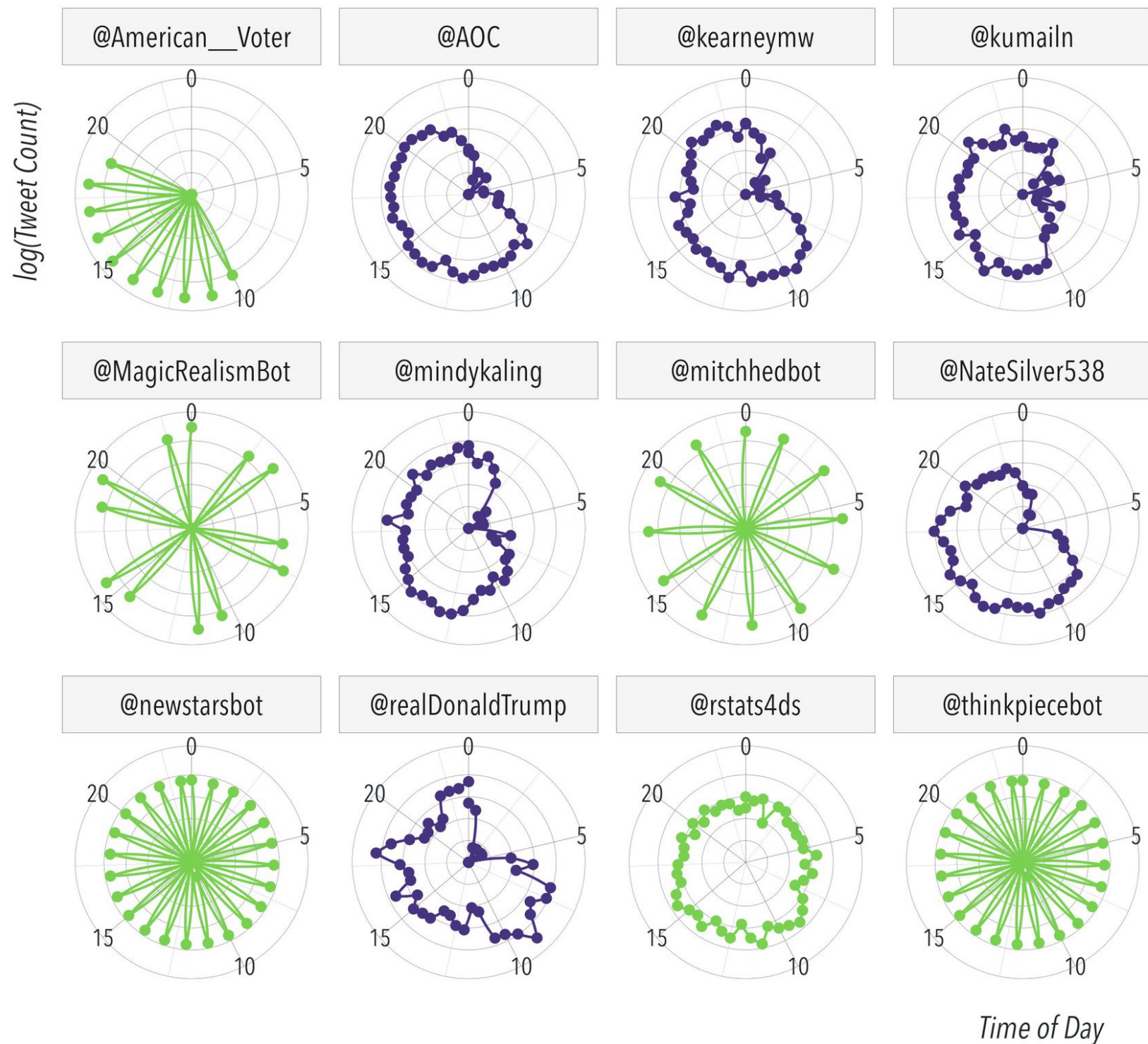
The AWEsomeE Process of Data Science!



What is Data Science?

Is it a TweetBot or Not?

Tweet frequency of accounts by time of day



Time of Day

Source: Michael Kearney

What is this course?

This course does not assume any background in data science! (But does assume some statistics and programming!)

While this is not a programming course, we will be doing a lot of it, and I will encourage good coding practice!

This course will pull from a wide array of data and topics to keep the fresh and exciting!

We will be using R as our computing language of choice but the concepts learned in this course translate to any language.

What is R?

R is an implementation of the S programming language, which was created 1976 by John Chambers at Bell Labs

R was created by Ross Ihaka and Robert Gentleman in 2000

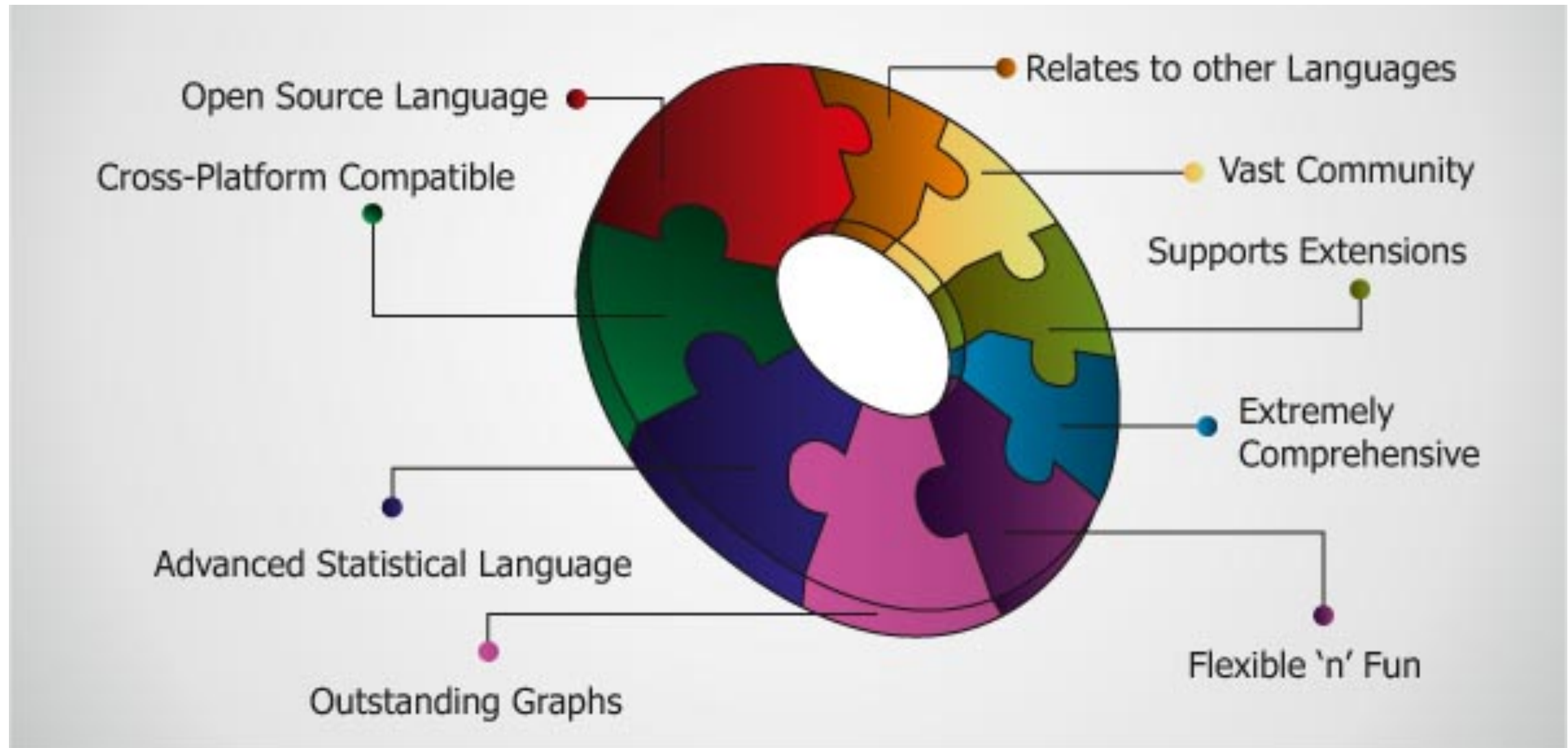
Although R and S are slightly different, they share the same design principle:

“[W]e wanted users to be able to begin in an interactive environment, where they did not consciously think of themselves as programming. Then as their needs became clearer and their sophistication increased, they should be able to slide gradually into programming, when the language and system aspects would become more important.”

R's basic rules:

- Everything that exists is an object
- Everything that happens is a function call

Why Learn R?



Let's Get Started!

Create a GitHub account: Go to github.com

Tips for selecting a username:

- Incorporate your actual name
- Reuse username from other contexts (Twitter, etc.)
- Pick a username you'll be comfortable revealing to your future boss
- Shorter is better than longer
- Be unique as possible in as few characters as possible
- Make it timeless. Don't highlight your current university, etc.
- Avoid words laden with special meaning in programming (ex. NA)

Source: [Happy git with R](#) by Jenny Bryan

Once you have created an account, send me an email with your username! (We have an organization that I will add you to!)

DataCamp

Go to:

https://www.datacamp.com/groups/shared_links/5373c2ec18020b107e519bc2271fc121ee7d2959

This link should give you access to our DataCamp site!

Make sure that you are in the group: MAT 219 Data Science I

RStudio Cloud

Go to https://rstudio.cloud/spaces/33434/join?access_code=zlwyKS1ywD6wzOoL19AEzFfdGsKP%2BhPoiYh0tBox and log in with your GitHub credentials!

RStudio Cloud is a web-based service that will hopefully house all of your work!

We'll talk about the features and workflow in RStudio Cloud later!

Doing Some Data Science!

United Nations Voting Analysis:

- Log on to RStudio Cloud and click on this course's workspace.
- Make a copy of the project for this application exercise and launch it.
- In the Files pane in the bottom right corner, spot the file called ``unvotes.Rmd``. Open it, and then click on the "Knit" button.
- Go back to the file and change your name on top (in the ``yaml`` -- we'll talk about what this means later) and knit again.
- Then, change the country names to those you're interested in. Your spelling and capitalization should match how the countries appear in the data, so take a peek at the Appendix to confirm spelling. Knit again. And voila, your first data visualization!