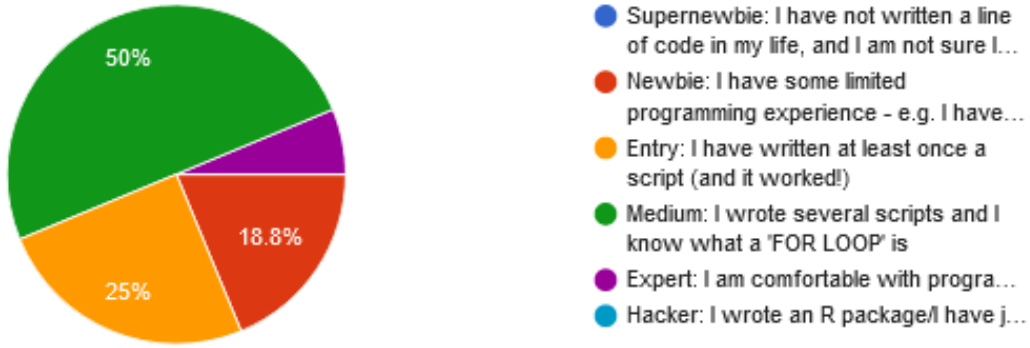
AGENDA

1. The skill level in the class
2. Data science workflow
3. Overview of tutorials
4. Programming basics
5. Exercises

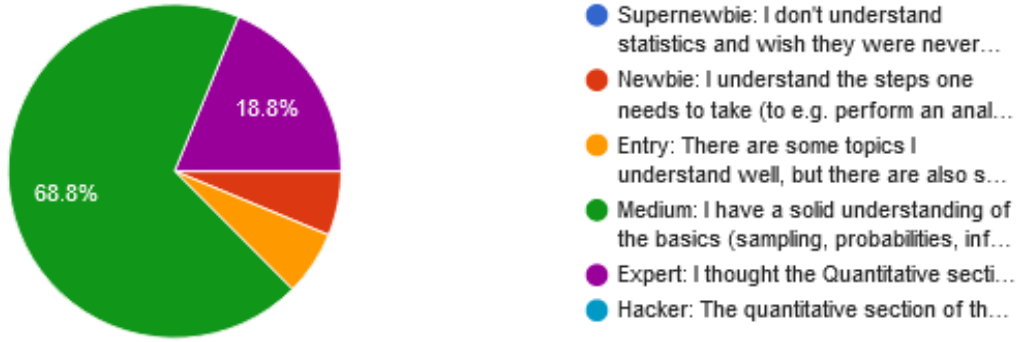
My programming proficiency:

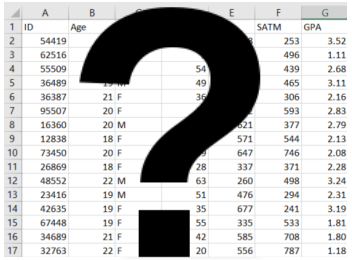
16 responses

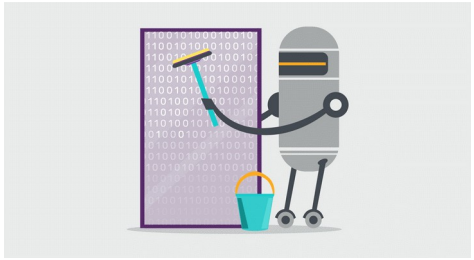


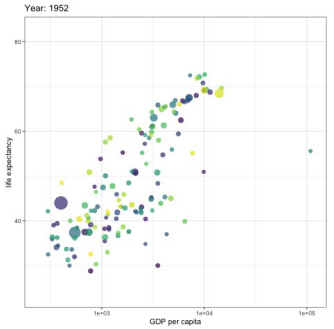
My statistics proficiency

16 responses



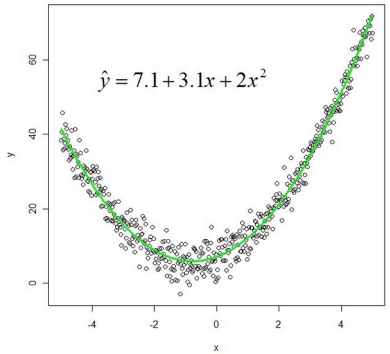






Descriptive Statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
price	74	6165.257	2949.496	3291	15906
mpg	74	21.297	5.786	12	41
rep78	69	3.406	.99	1	5
headroom	74	2.993	.846	.846	5
trunk	74	13.757	4.277	5	23
weight	74	3019.459	777.194	1760	4840
length	74	187.932	22.266	142	233
turn	74	39.649	4.399	31	51
displacement	74	197.297	91.837	79	425
gear_ratio	74	3.015	.456	2.19	3.89
foreign	74	.297	.46	0	1



OVERVIEW OVER TUTORIALS

1. Basic programming pt. 1
2. Basic programming pt. 2
3. Data wrangling
4. Data visualization
5. Exploratory data analysis
6. Supervised machine learning
7. Unsupervised machine learning

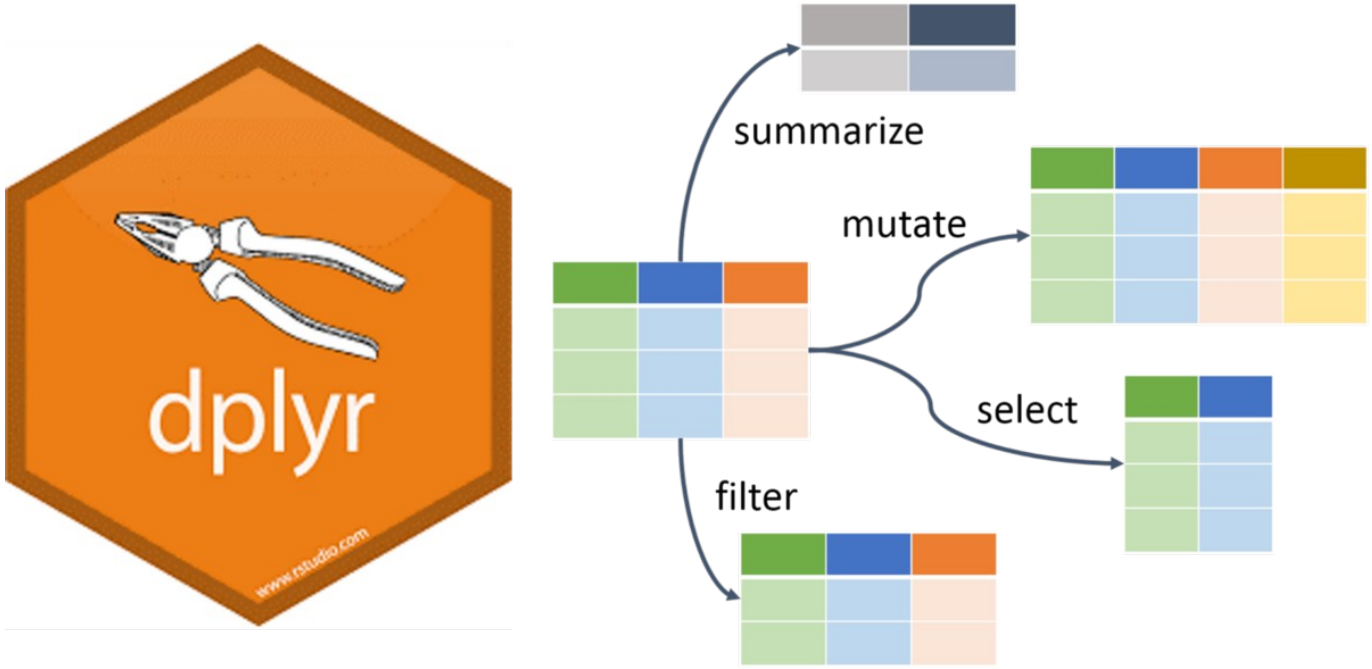
8. Text mining (Michael Zaggl)
9. Guest talk
10. + 11. Exam work

TUTORIAL 1: BASIC PROGRAMMING

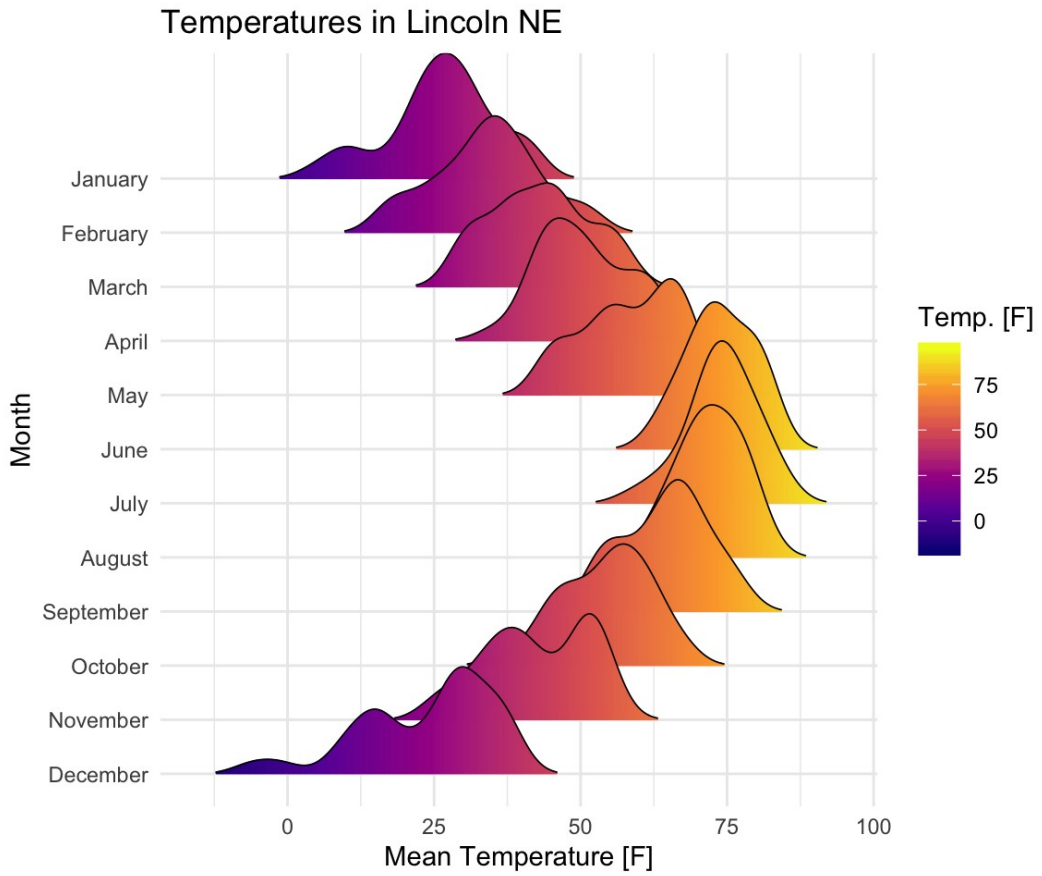
TUTORIAL 2: BASIC PROGRAMMING

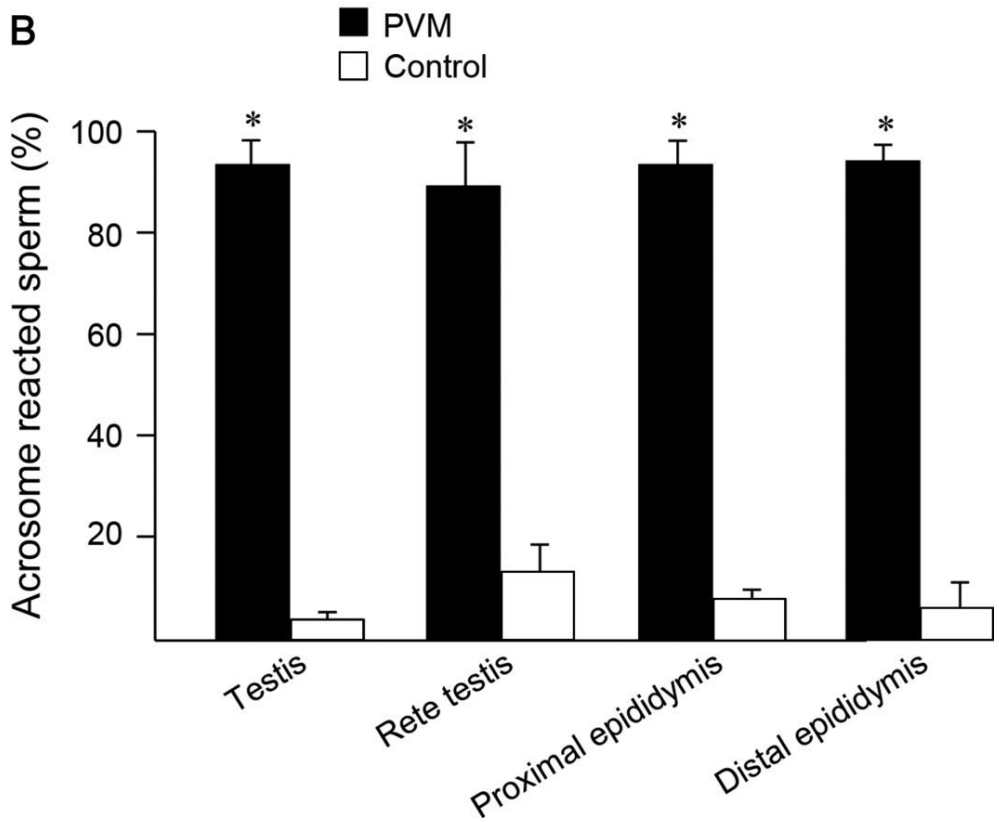
Work with actual datasets in R

TUTORIAL 3: DATA WRANGLING

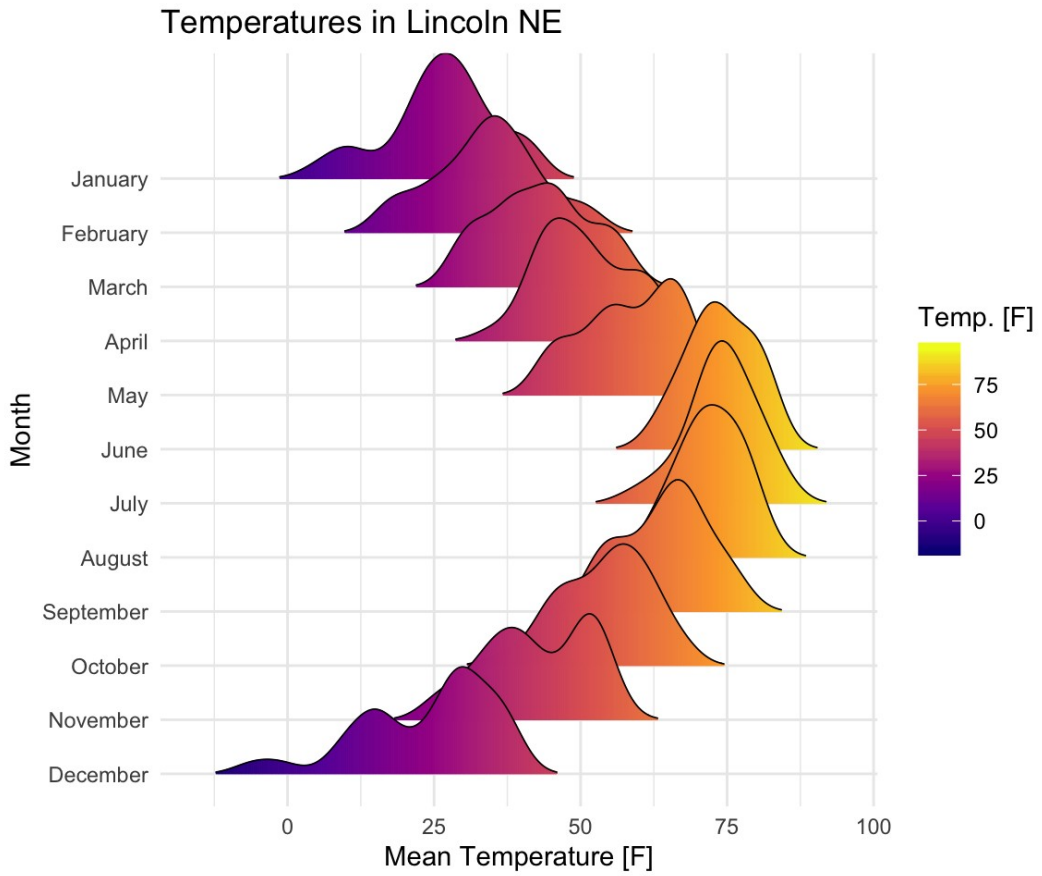


TUTORIAL 4: DATA VIZ

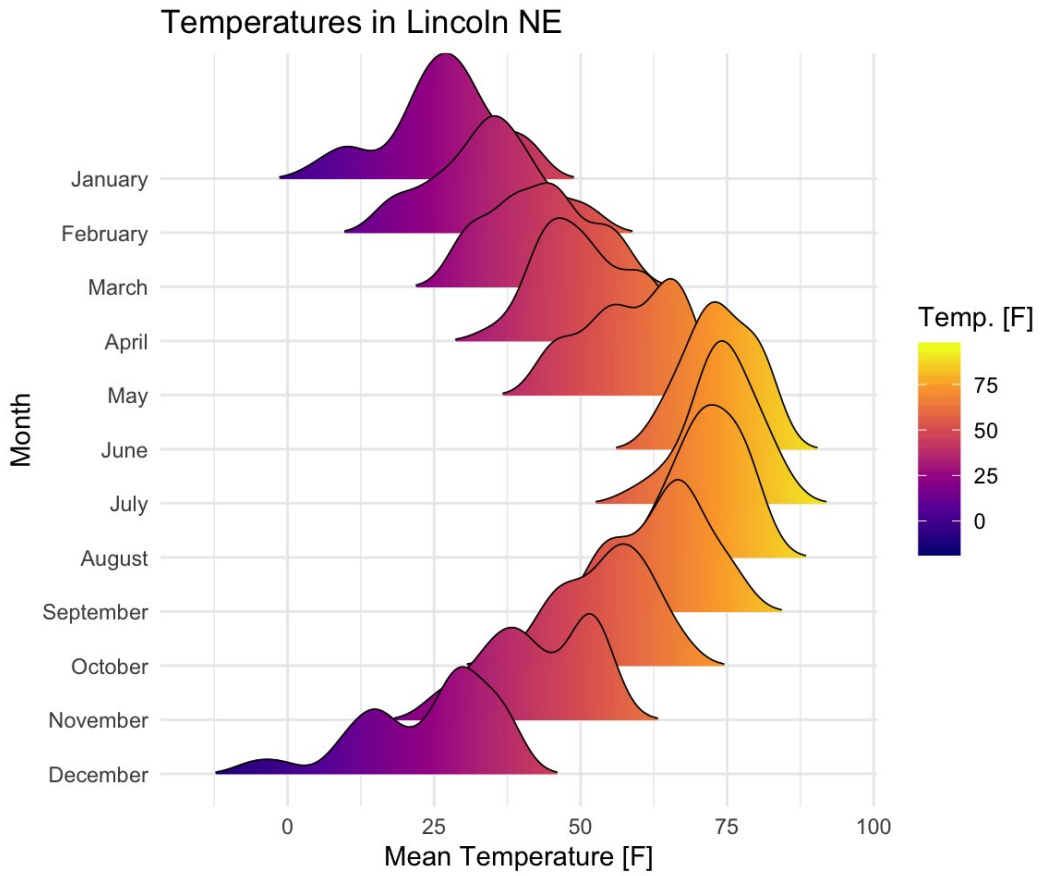


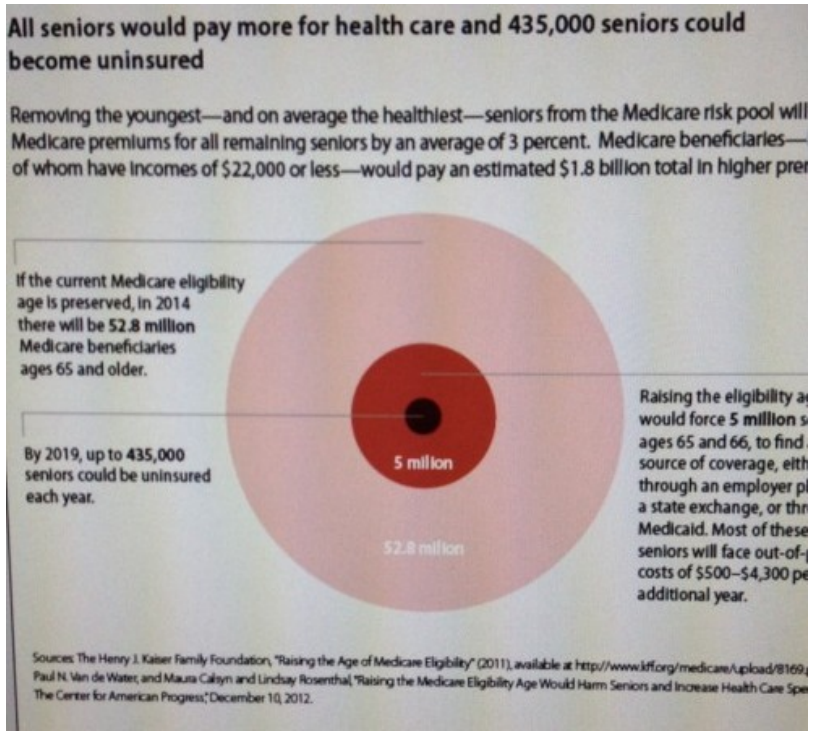


TUTORIAL 4: DATA VIZ



TUTORIAL 4: DATA VIZ



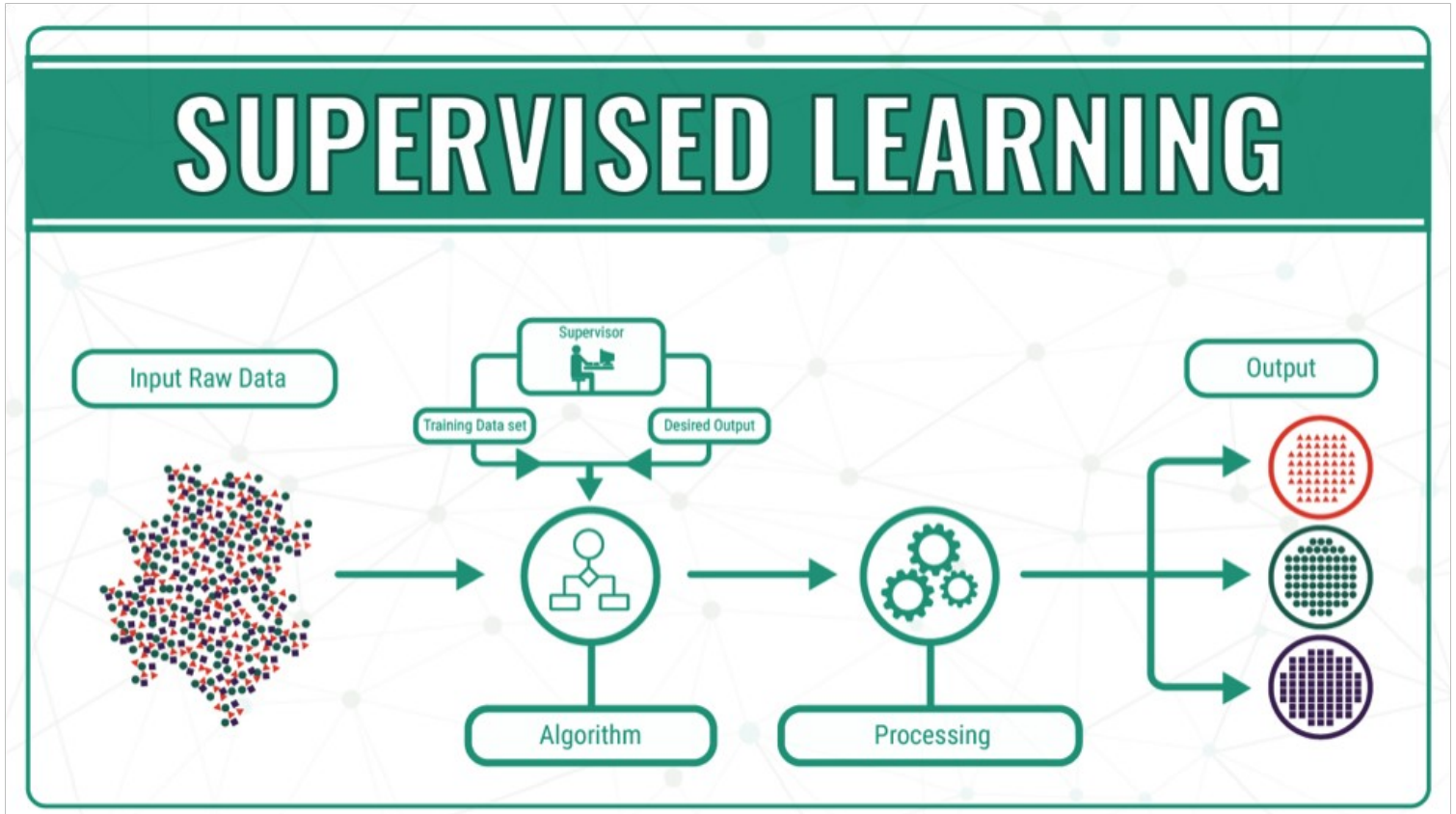


https://accidentalboobcharts.tumblr.com

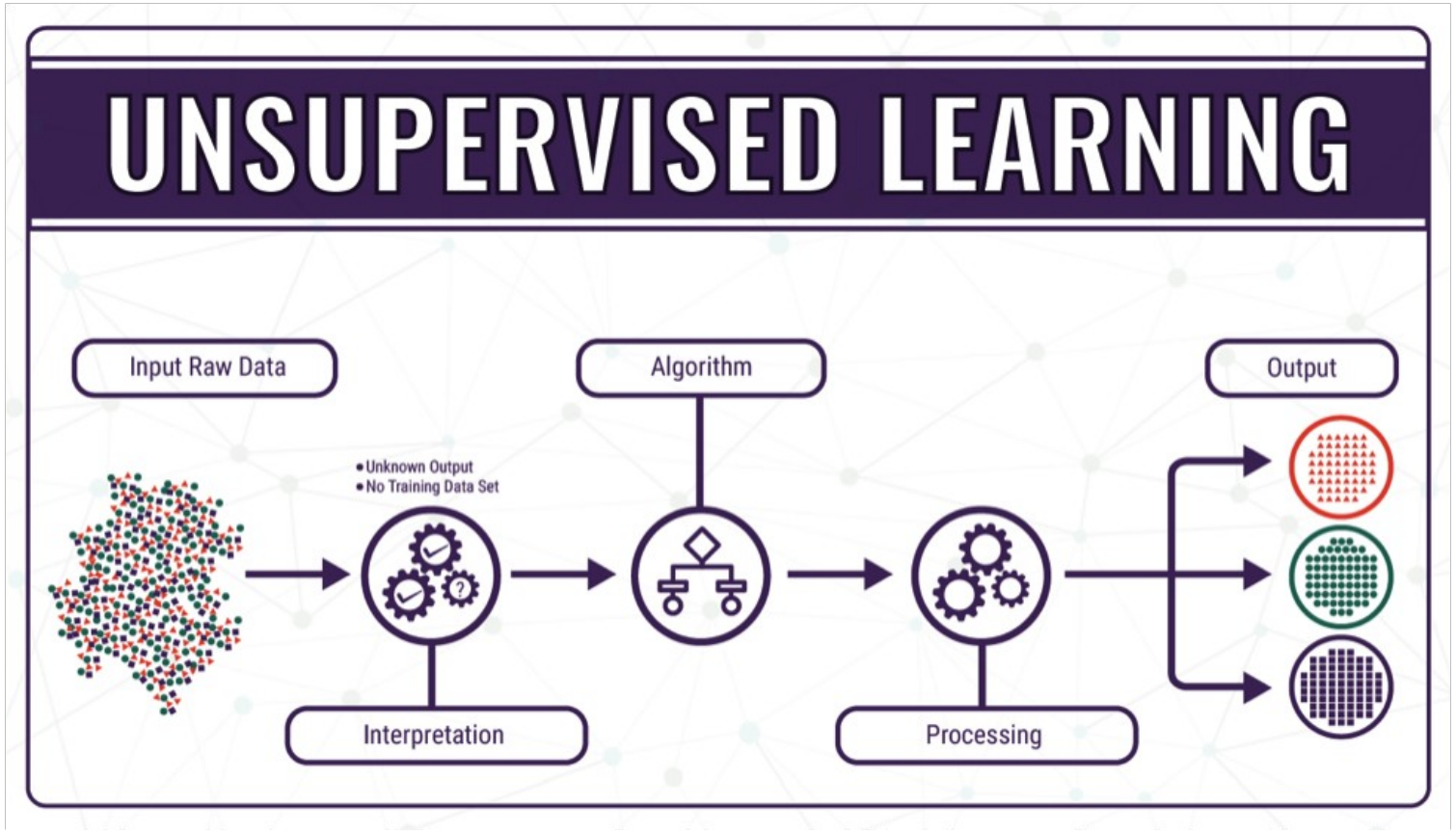
TUTORIAL 5: EXPLORATORY DATA ANALYSIS

CAROLINE KJÆR BØRSTING
PHD STUDENT

TUTORIAL 6: SUPERVISED MACHINE LEARNING



TUTORIAL 7: UNSUPERVISED MACHINE LEARNING



PROGRAMMING BASICS

Variables – how data is represented, e.g. Age, Gender, but also more complex structures such as tables, or, as you will see data frames

Loops – allow you to do the same thing repeatedly

Conditionals – if <condition> do <command>. So e.g. if age is above >18, allow person to buy liquor

Input/output – these are the things that allow your computer to interact with "the real world" – in our case, how to load data and you to export e.g. tables, figures etc.

Subroutines/functions – chunks of commands that can be reused.

Commands – this is just fancy programming language for actions. E.g. print, modify value, assign value, delete.

UNDERSTANDING ‘FUNCTIONS’ AND ’PACKAGES’ FOR NON-PROGRAMMERS

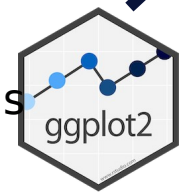
Eat, sleep, shit, repeat

Take on different jobs

Celebrate the seasons

R Core team

Import data
Mean of variable
Plot ugly histograms



Cool plots

Text mining

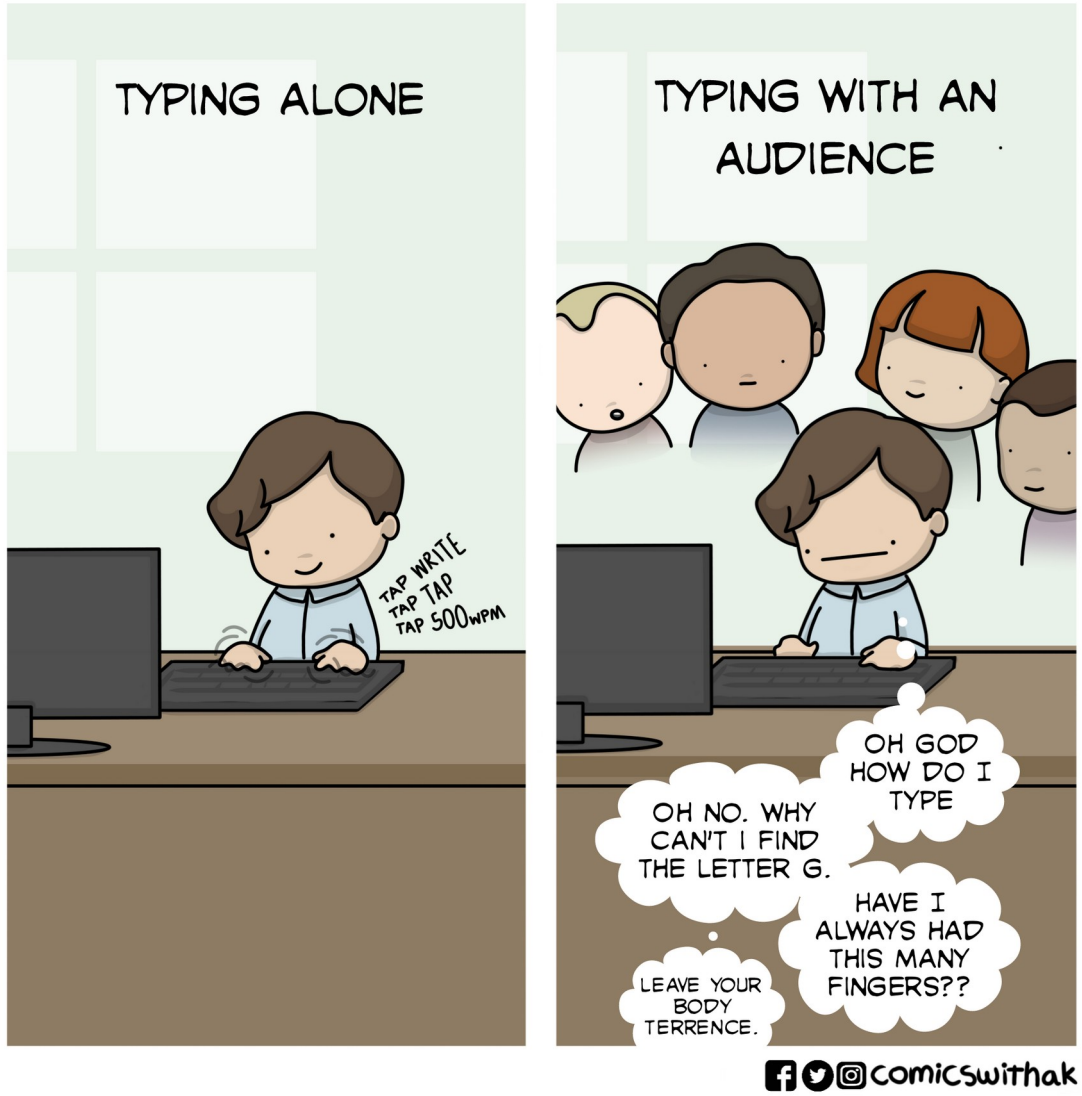
R EXAMPLE

Console (write and execute code)

Global environment (imported datasets, saved variables)

Tabs: files, plots, packages, help, viewer

Script


TYPING ALONE
TAP WRITE TAP TAP TAP 500WPM
TYPING WITH AN AUDIENCE
OH GOD HOW DO I TYPE
OH NO. WHY CAN'T I FIND THE LETTER G.
HAVE I ALWAYS HAD THIS MANY FINGERS??
LEAVE YOUR BODY TERRENCE.
comicswithak

SWIRL

Learn R, in R.

swirl teaches you R programming and data science interactively, at your own pace, and right in the R console!

```
 1: Basic Building Blocks      2: Workspace and Files       3: Sequences of Numbers      4: Vectors                  5: Missing Values          6: Subsetting Vectors       7: Matrices and Data Frames
 8: Logic                      9: Functions                10: lapply and sapply        11: vapply and tapply       12: Looking at Data        13: Simulation              14: Dates and Times
15: Base Graphics
```


Doctors: Googling stuff online does not make you a doctor.
Programmers:
STARECAT.COM

https://twitter.com/rstats4ds

https://twitter.com/rfunctionaday

https://twitter.com/RLadiesGlobal

NEXT TIME

- Date: 16.09.2021
- Before class: finish swirl tutorial (module 1,3,4,5,6,7)
- Topic: Basic programming
- Exercise: basic programming steps on actual dataset in Rstudio (+ knitting Rmarkdown to HTML)

DEPARTMENT OF MANAGEMENT
AARHUS UNIVERSITY