Weather & Corona

Michele Coscia

First Year Project #1

February 3rd, 2022

Outline

- Logistics
- Project Background
- Kickoff Q&A

Logistics

Skills for Project 1

- Geospatial data wrangling
- Choropeths
- Multiple hypothesis testing
- Multivariate regression

Work Organization: GIT

- Study well the Github tutorial from previous lecture
- Follow best practices
- Remember: code & docs, no data (use .gitignore)

Project Folder Suggestion

- Data
  - Raw (Original, immutable)
  - Interim (Intermediate datasets)
  - Processed (Final datasets for the analysis)
- Notebooks (Polished “start to finish” code for report)
- References (Data documentation, manuals, literature)
- Reports (Your outputs for the project: the final handout, slides, etc)
  - Figures
- Code (Working directory with all the code organized in logical subfolders)

Exercises

- Doing the exercise = working on the project
- The tasks will help you with your report
- Follow them to get the basics down so you have more time for the interesting stuff

Hand-in

- gitlog.txt
- code.zip
  - No data, unless it’s 3rd party
- report.pdf
  - Formatting instructions on LearnIT

Report

1) Introduction (context and motivation)
2) Data (with cleaning steps, including missing data)
3) Results & Discussion
4) Limitations (short-coming(s) of your methodology/data)
5) Concluding remarks and future work
6) Disclosure statement (optional)

Tasks

0) Data filtering & cleaning
1) Single variable analysis
2) Associations
3) Map visualization
4) Open question

Exam

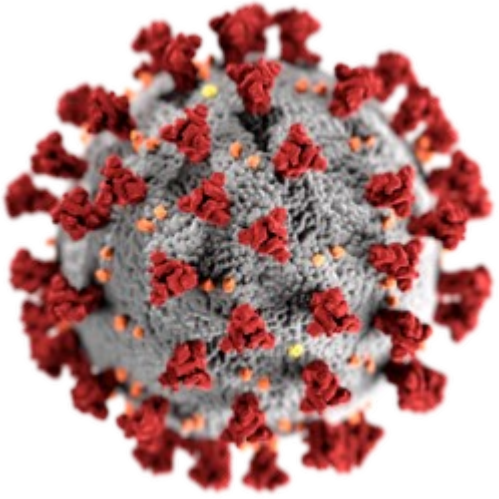
- 30 minutes per group
- Taken as a group all together
- Project Presentation
- Questions: on project AND on theory
- March 8th & 10th
- Only feedback, no intermediate points (sorry)

Lecture Plan

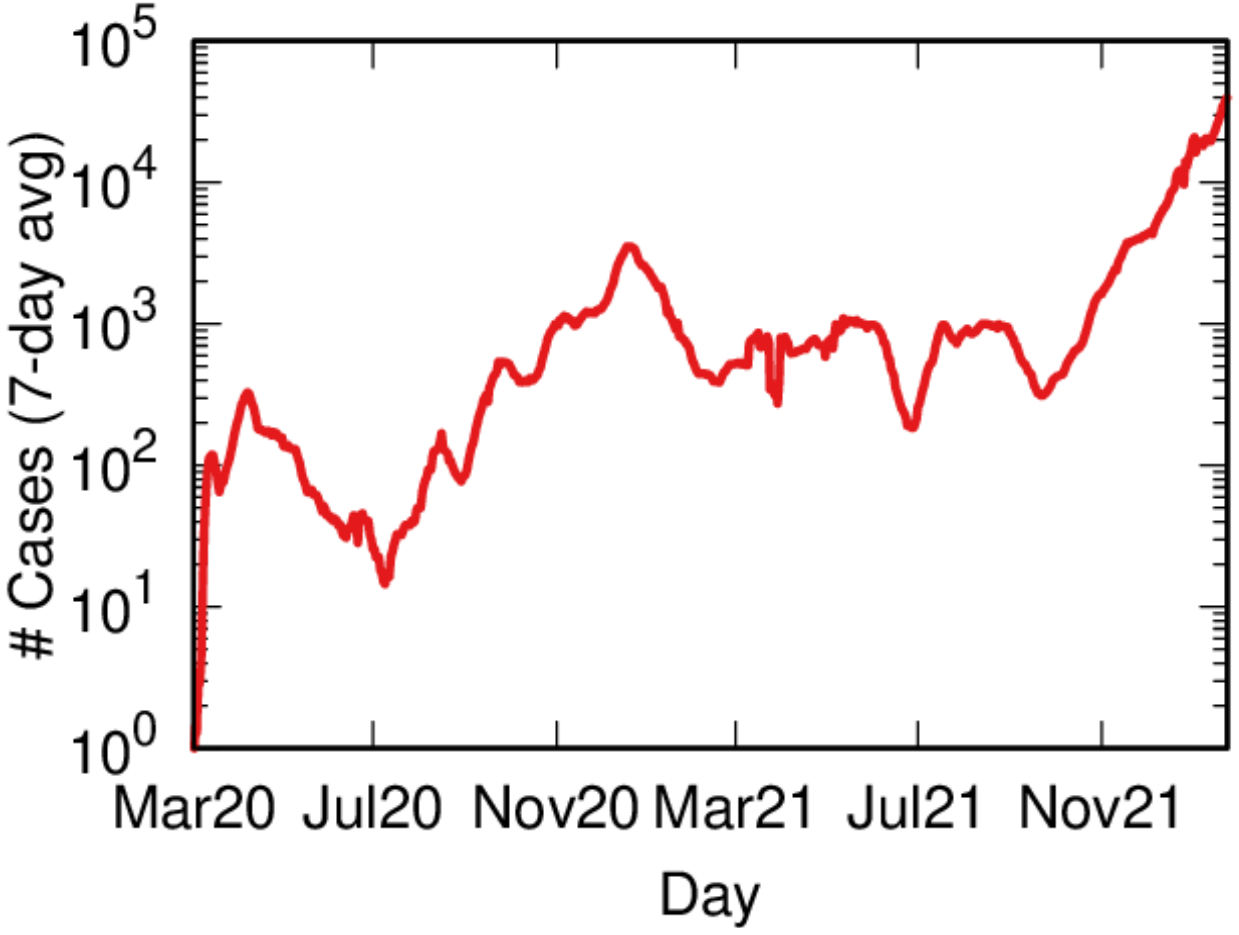
1) **(Today) Intro**
2) (February ~~10th~~ 8th) Geospatial Basics
3) (February 15th) Estimating Associations
4) (February 17th) Multivariate Regression
5) (February 22nd) Interventions
6) (February 24th) Project Run Through
7) (March 1st) Q&A – Open Supervision
8) (March 3rd) Q&A – Open Supervision

Project Background

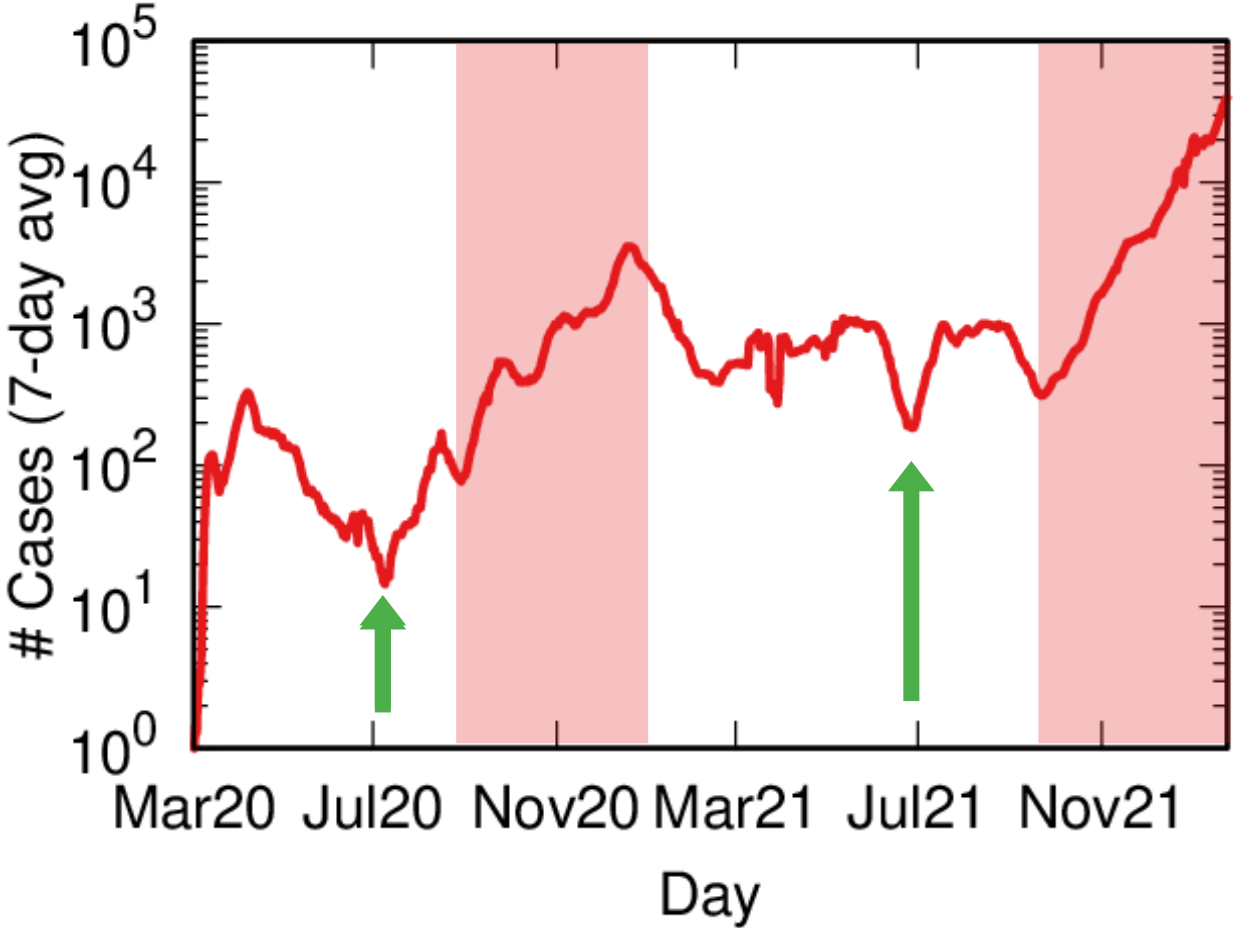
No Need to Introduce you



Huge Disruptions



Mmmh…


10^5
10^4
10^3
10^2
10^1
10^0
Cases (7-day avg)
Mar20
Jul20
Nov20
Mar21
Jul21
Nov21
Day

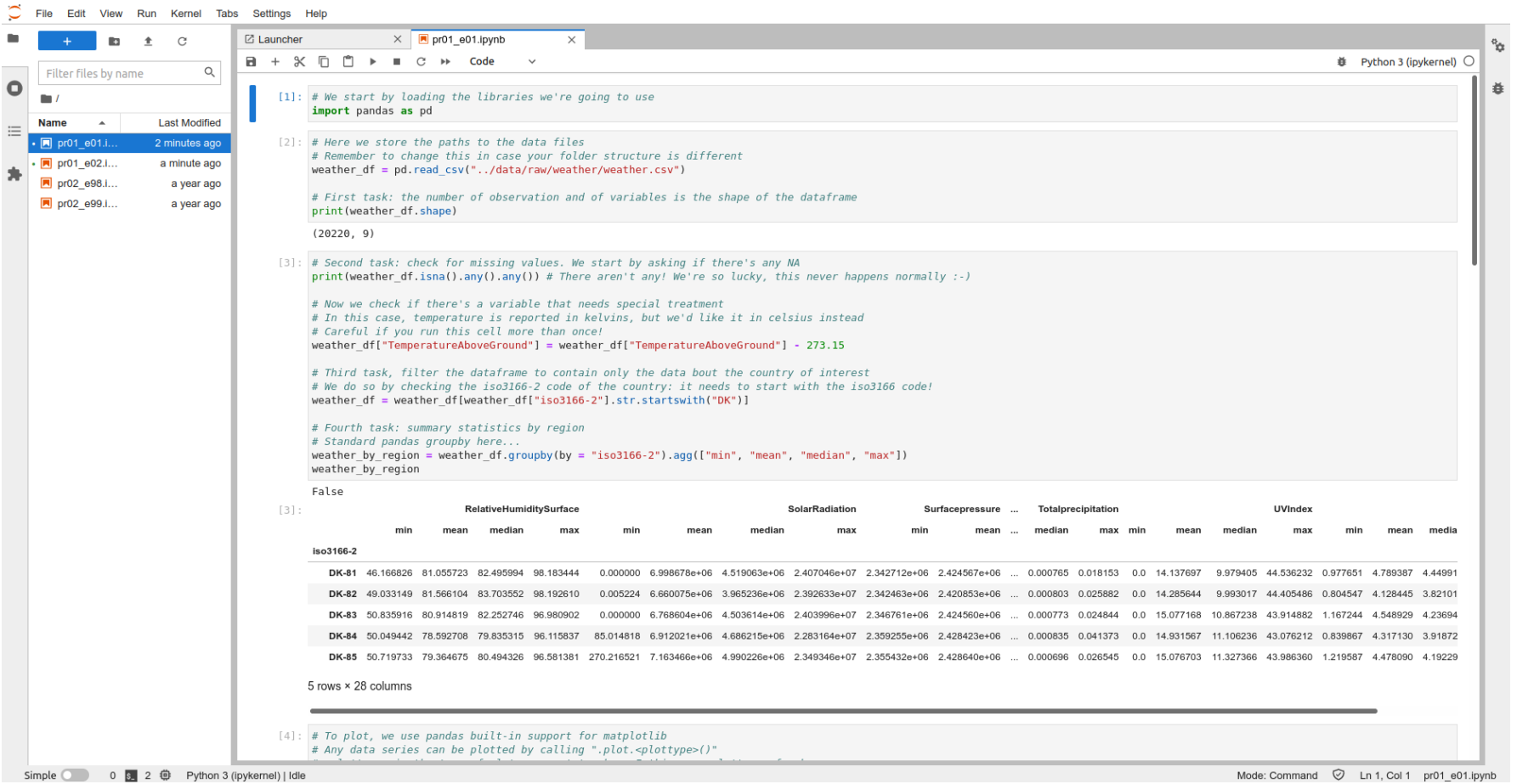
Seasonal Diseases

- Cold
- Flu
- Covid?

Let’s discuss:
what could be the cause?

Congratulations, you have a working hypothesis for your task 4 :-)

Quick Look at Weather Data



Q&A