

Math 189: Introduction

I. Overview

Syllabus

Instructor goes through syllabus.

What is R?

R is a computer language for doing statistical analysis. It is command-line style, so no compilation is needed before running (unlike C, Fortran, etc.). R is free, and you should install it on your PC. Your TA (as well as the instructor) can help with R; also see this tutorials: *An Introduction to R*. <https://cran.r-project.org/doc/manuals/R-intro.pdf>

What is RStudio?

RStudio is an IDE (Integrated Development Environment), which allows you to write and execute R code, participate in projects, view output, and more. This is free; please download and install.

What is Python?

Python is also a command-line style language, and can be downloaded for free. For production-quality code, it is superior to R. PyCharm is a Python IDE. The class will be taught in R.

What is R Markdown?

Markdown is a *plain text formatting syntax*. You will use this to generate reports for your group homeworks and final project. R Markdown extends Markdown by embedding R code.

What is GitHub?

Course material will be on <https://github.com/tuckermcelroy/ma189> (as well as Canvas). This is a *repository* where you can view files (like Markdown files and data sets) and download them to your PC. Click *Raw* to view underlying R Markdown source file.

First Tasks

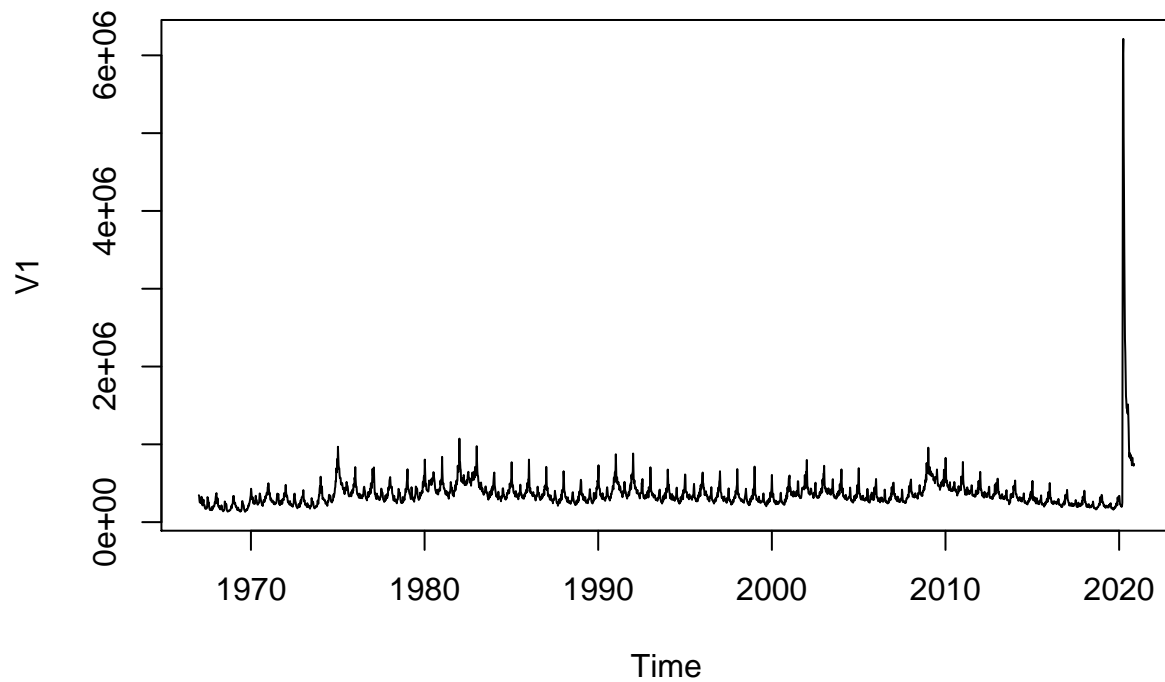
Install R and RStudio, and get familiar with accessing files on Canvas and GitHub. Review the background reading:

1. Writing Data Analysis Reports: *WritingReports.md*
2. Some slides by Jonathan Taylor at Stanford Statistics Department, giving a constrast of *Statistics* and *Data Science*. <https://math189.edublogs.org/files/2018/01/Statistics-and-Data-Science-1vh53fi-1ylpezs.pdf>

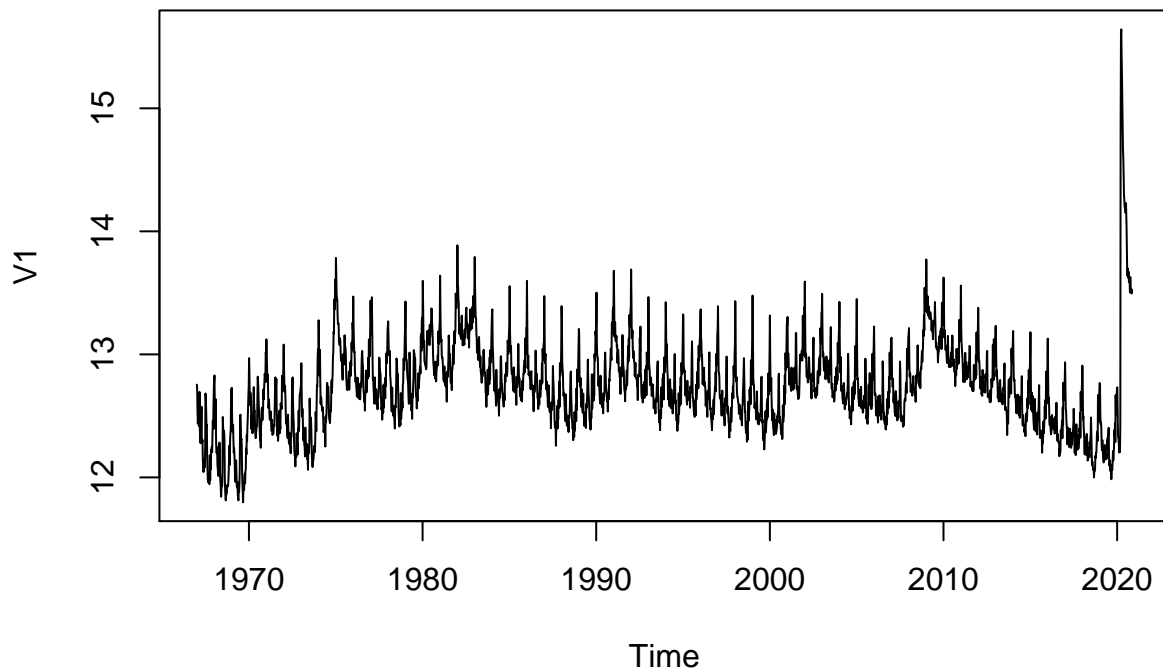
II. National Unemployment Insurance Weekly Claims Data

Unemployment Claims Surge due to Covid-19

```
claims <- read.table("C:\\Users\\neide\\Documents\\GitHub\\ma189\\Data\\InitialClaims.dat")
period <- 365.25/7
claims <- ts(claims, start = c(1967,1), frequency = period)
plot(claims)
```



```
plot(log(claims))
```



Exploratory Data Analysis

First examine your data.

1. Catch mistakes in how data was entered
2. Discern patterns in the data
3. Check model assumptions
4. Generate working hypotheses

Prior Knowledge

New plague strikes the world with vast repercussions to health and the economy. Unemployment in certain sectors rises in response to government lockdowns.

Paradata (Facts About the Data)

Data collected by Bureau of Labor Statistics: <https://oui.doleta.gov/unemploy/claims.asp>

“The Unemployment Insurance weekly claims data are used in current economic analysis of unemployment trends in the nation, and in each state. Initial claims measure emerging unemployment and continued weeks claimed measure the number of persons claiming unemployment benefits.”

Covers Week 1 of 1967 through week 46 of 2020.

What Questions to Ask

- How are claims impacted by the pandemic?

- When did the effect begin? Has it ended?
- How do we model this effect?