

# Class 10: Data wrangling III

---

February 22, 2018



# General

# Announcements

- Reading for next class: *R for Data Science*
  - All of **chapter 7**
- Homework 1 due Friday, February 23rd by 11:59pm
- Homework 2 will be posted on Friday as well, will be due on Friday, March 9th by 11:59pm (Friday before Spring Break)
- Schedule adjustments
  - Midterm projects will be presented on March 22, the week **after** Spring Break .
  - Rest of course calendar will be updated to reflect this change and our current pacing

# Group Activity

# Group activity instructions

The goal of this group exercise is to answer the following question about the

`nycflights13` dataset:

What month had the highest proportion of cancelled flights?  
What month had the lowest?  
Interpret any seasonal patterns that you see.



# Group activity instructions

The goal of this group exercise is to answer the following question about the

`nycflights13` dataset:

What month had the highest proportion of cancelled flights?  
What month had the lowest?  
Interpret any seasonal patterns that you see.



1. Form groups of 2 – 5 students and choose a scribe
2. The scribe is to download the `nycflights13_exercise.Rmd` file I posted on Slack and open it in RStudio
3. Work through the sections of the RMarkdown file as a group, which is designed to lead you through the necessary steps for answering the above question
4. The scribe will write down your solutions for each section
5. When you're done, the scribe should attach the completed RMarkdown file to a Slack Direct Message **addressed to Dr. Glasbrenner and the other group members**