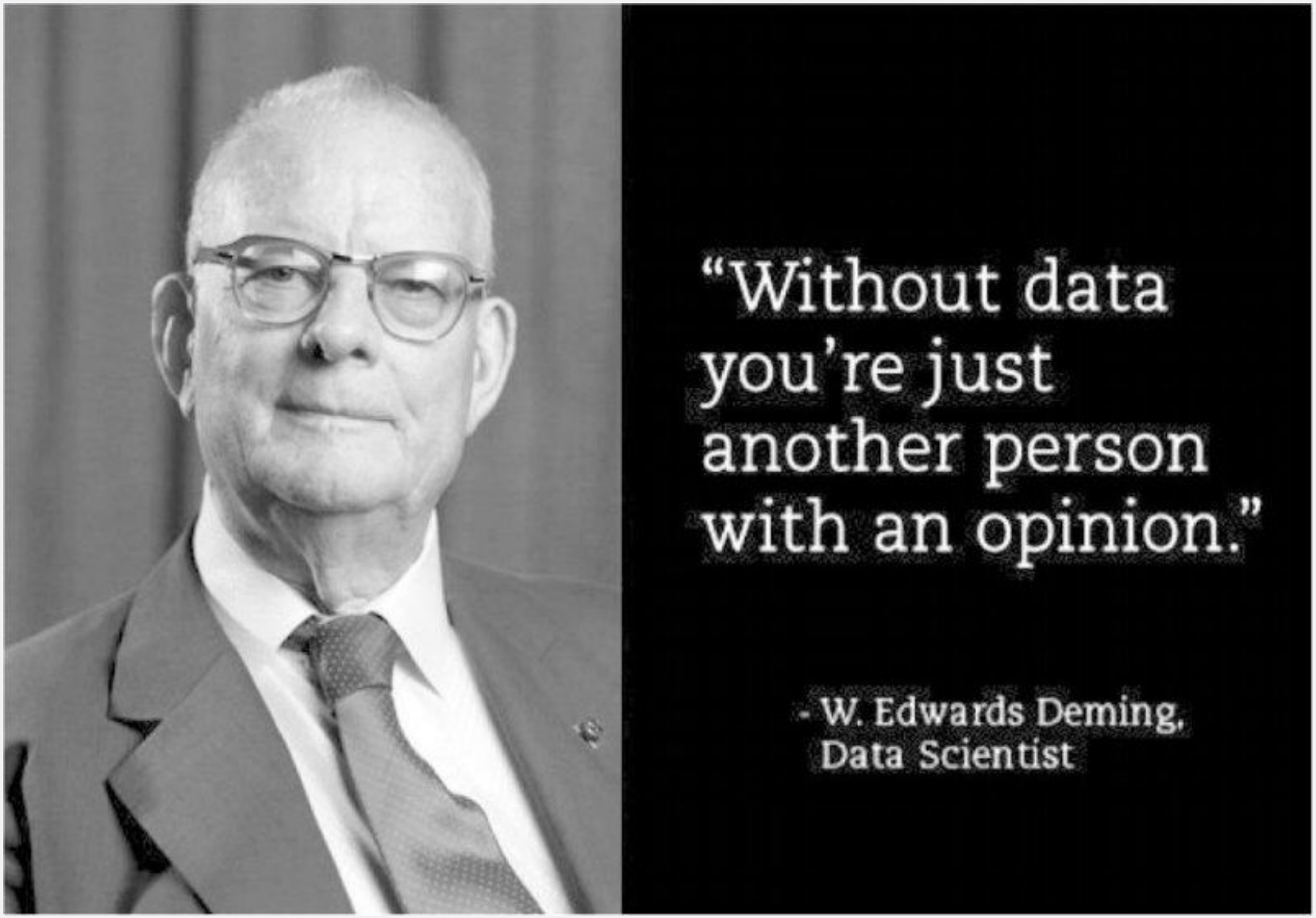
PSC 400

SYRACUSE UNIVERSITY

DATA ANALYTICS FOR POLITICAL SCIENCE

INTRODUCTION

DATA ANALYSIS



WHAT IS THIS CLASS ABOUT?

- Learn how to conduct data analysis yourself
- Finding data, data cleaning and data manipulation, data visualization, and data analysis

DATA ANALYSIS

Forbes

Data Scientist Leads 50 Best Jobs In America For 2019 According To Glassdoor

- Data Scientist is the best job in America for the 4th year in a row (4.7 job score, 4.3 job satisfaction rating) with 6,510 open positions paying a median base salary of $108,000.

DATA ANALYSIS

The New York Times

Data Analysts Captivated by R's Power

R is also the name of a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca partly because data mining has entered a golden age, whether being used to set ad prices, find new drugs more quickly or fine-tune financial models. Companies as diverse as Google, Pfizer, Merck, Bank of America, the InterContinental Hotels Group and Shell use it.

But R has also quickly found a following because statisticians, engineers and scientists without computer programming skills find it easy to use.

"R is really important to the point that it's hard to overvalue it," said Daryl Pregibon, a research scientist at Google, which uses the software widely. "It allows statisticians to do very intricate and complicated analyses without knowing the blood and guts of computing systems."

CLASS WEBSITE

- http://simonweschle.com/psc400

- Refresh using
  - Windows: ctrl + F5
  - Mac/Apple: Apple + R or command + R

APPROACH

- Work through assigned reading *before* class
- Class
- After class: review reading if necessary
- Short review exercises (approx. weekly)
- Problem sets (every other week)
- Your own data analysis project (over course of semester)

ASSIGNMENTS/GRADES

- Class Participation
- Programming Review Exercises
  - (roughly) weekly, graded P/F
- Problem Sets
  - 5-6 total, every 2 weeks
- Data Analysis Memos
  - 2-3 pages, 5-6 total, every 2 weeks
- Data Analysis Paper
  - Due May 10

DATA ANALYSIS PAPER

- Answer your own research question
  - Memos: your research question, your data, data cleaning, descriptive statistics, bivariate relations, (first) regression results
  - Final paper: set out your research question, explain the data and statistical methods you use to investigate it, and describe what, based on your data analysis, the answer is

CLASS TOPICS

- Getting Started with R
- Causality with randomized experiments
- Single Variables
- Finding and Cleaning Data
- Bivariate Relationships
- Multivariate Regression
- Spatial Data, Network Data, Text as Data, Website Scraping
  - will choose some of those topics based on student interest
- Data Analysis Paper Workshop

COVID

- In-person as long as it’s safe to do so
- If you are sick (even just a runny nose), do not attend in-person. Email me and I’ll turn on Zoom

COVID, WHAT ELSE?

- If you experience any hardship (Covid-related or not) that interferes with your academics, *please* get in touch with me
- I'll work with you. There's always something that can be done. But I need to know about it
- You do not have to reveal personal details to me
- Talk to me as early as possible, harder to do something after the fact
- If you ask yourself whether you should talk to me, talk to me

TEXTBOOK

- Elena Llaudet and Kosuke Imai: Data Analysis for Social Science, A Friendly Introduction
  - Link on Blackboard (please do not share with anyone)
- Other readings on Webpage/Blackboard

STUDENT HOURS

- Monday 11:00-1:00
  - 530 Eggers or Zoom (info on syllabus)
  - Or by appointment (email me)
  - Your chance to talk to me about course material, college in general, or just to chat
  - *Please* talk to me as early as possible about any issues or questions
    - If I know what's going on, I can almost always help in some way (or I know someone who can help)
    - It's much more difficult to do something after the fact

DICTIONARY

- Create a folder called PSC400 (or something like that)
  - on Desktop or (preferably) somewhere else
- This is where you'll save everything from this class (code, datasets, etc)

R AND RSTUDIO