

A close-up photograph of two hands, likely belonging to a man, against a dark background. The hands are clenched into loose fists. Resting on the palm of the left hand is a single, shiny red capsule. Resting on the palm of the right hand is a single, shiny green capsule. The lighting is dramatic, highlighting the texture of the skin and the metallic sheen of the capsules.

JANUARY 16, 2018

---

# INTRO TO PS239T

## MOTIVATION

---

# WHY ARE WE HERE?



# THE DREAM

- ▶ There is a lot **more** data available to researchers now than ever before [thanks, internet]
- ▶ The data is a lot easier to **find**, **acquire**, and **store** than ever before [thanks, computers]
- ▶ The data is more **varied**, and in many cases, **higher quality**, than ever before [thanks, science]



# THE REALITY

- ▶ “Every day, scholars all over the world waste time wrestling with computers. Tasks that should take a few moments take hours or days, and many things never work at all. When scholars try to get help, they are inundated with unhelpful information, and give up.” - Rochelle Terman, creator of PS239T
  
- ▶ “Some other time”
- ▶ Too many options
- ▶ Inefficient to learn on your own
- ▶ Shaming



# SO, WHY BOTHER?

- ▶ “Some other time”
  - ▶ Code is reusable, accumulating labor (aka, capital).
- ▶ Too many options
  - ▶ Learn the logic now, specialize later.
- ▶ Inefficient to learn on your own
  - ▶ Totally true. Do it here.
- ▶ Shaming
  - ▶ Laugh all the way to the bank (or the World Bank).
- ▶ Plus...
  - ▶ Better scholarship.



## OBJECTIVES

---

## MY PITCH

Coding is like learning any other language

# AIMS OF THE COURSE

- ▶ **Skills** (the grammar)
  - ▶ Bash, R, HTML/CSS, Python
  - ▶ Transfers between many languages
  - ▶ Relies on shared logics and structures
  - ▶ Constrains what can and can't be done
  - ▶ Used to **make sense** of things: shaping, analyzing, plotting
- ▶ **Applications** (the vocabulary)
  - ▶ APIs, webscraping, packages in R and Python, etc.
  - ▶ Used to **observe** and **obtain** things—in our case, data

# AIMS OF THE COURSE

- ▶ ***Learn to learn***
- ▶ Expect to learn how to:
  - ▶ Navigate & operate effectively in a UNIX environment
  - ▶ Master basic Git and GitHub workflows
  - ▶ Write, execute, and debug R code for data cleaning, statistical analysis, data visualization, and machine learning
  - ▶ Parse HTML, CSS, and Javascript for the purposes of using tools like APIs, webscraping, and Qualtrics
  - ▶ Write, execute, and debug Python code for text analysis, as well as other computing tasks
  - ▶ Find answers to things we won't cover in this class

# (NOT THE) AIMS OF THE COURSE

- ▶ What won't you learn?
  - ▶ Professional programming or software development
  - ▶ Formal statistics or computer science
  - ▶ All the answers
    - ▶ And that's okay...coding wizardry is **not** about knowing everything ahead of time. It's about being able to **adapt** to challenges when you encounter them.



# MEASURING CANDIDATE GENDER

### ▶ The task:

- ▶ Read a bunch of voting pamphlets (PDFs) to get candidate images and information; turn the latter into a dataset (CSV)

### ▶ Estimated time:

- ▶ 16 elections (and counting) ×
- ▶ 800 pages of PDFs per year ×
- ▶ 2 candidates a page ×
- ▶ 10 mins a page to get pics and parse info =
- ▶ 1.1 YEARS coding eight hours a day, five days a week, just to get the data
- ▶ Not even accounting for fatigue and human error

8 Candidates | Republican Candidates

United States Senator



**Mark Callahan**

Republican

**Occupation:** Information Technology (IT) Consultant  
**Occupational Background:** IT industry consulting/permanent positions, at local/national companies, over 13 years.

**Educational Background:** B.S. Business Administration / MIS, e-Law Minor, OSU – Corvallis, OR; MCP, MCSE Training Program, THE COMPUTER INSTITUTE – Rockville, MD  
**Prior Governmental Experience:** Marion County Republican Party PCP, 2013 – Present; Lane County Republican PCP, 2011-13; Eugene Income Tax Opposition Statement Committee, 11

**Leadership, Integrity, Character, Honor & Trust**  
**Address Government Infringement:**

We must stop our government from violating our constitutional protections. From the NSA spying us, to themacking our 2nd Amendment rights, to the IRS targeting, to political correctness destroying our religious freedom. Our nation's sovereignty is under attack with the implementation of Agenda 21 and Common Core. As your constitutionally conservative Senator, I will stand strong, with Senator Ted Cruz for the rights of all Americans.

**conomy:**

As your Senator, who believes in the free market, I will work to make America's economy strong again, beginning with the Full Repeal of ObamaCare a.k.a "MerkleyCare".

**Balance Budget: Reduce Spending:**

As your fiscally conservative Senator, and a believer in limited government, I will reduce wasteful spending and push for a balanced budget amendment.

**Help Americans:**

We're at a crossroads affecting the very freedom our constitutional republic was founded upon. Supporting moderate candidates, that don't win general elections, is not an option. We need to elect a true conservative candidate that will fight for American values. We will win this election, by distancing ourselves from career politicians who are only looking to advance their own careers. We will win with a Pro-Life candidate that embraces our conservative values. A candidate not afraid to hold the President accountable for lying to us, without hesitating to convict him, if the House impeaches him. A candidate that isn't afraid to take on the establishment, promote freedom from government intrusion, dependency, poverty. My fellow Americans, I Am That Candidate. Thank you and May God Bless America.

Campaign website: [www.CallahanForOregon.com](http://www.CallahanForOregon.com)  
(This information furnished by Callahan For Oregon.)

United States Senator



**Timothy Crawley**

Republican

**Occupation:** Attorney  
**Occupational Background:** Nichols & Associate  
**Educational Background:** U.C. Hastings, J.D.; College, B.A.

**Prior Governmental Experience:** U.S. District Court, Oregon; Supreme Court of California; Federal Reserve

I was born and raised in Cottage Grove where my family still resides. I consider it a personal obligation to help maintain its natural beauty and a healthy ecology by continuing to enjoy the benefits of a free market, without monopolistic power. I have travelled to nearly thirty countries and explored a multitude of foreign cultures and environments, but Oregon has always remained my home. That is why, after completing my legal education in California and beginning my professional career there, I have returned to Oregon.

My family always encouraged me to serve others. I advocated for civil rights in the Middle East; crewed on an Alaskan fishing boat; educated the public for the Federal Reserve Bank; and represented tech entrepreneurs in Silicon Valley. Since my return to Oregon, I have been a volunteer in environmental restoration efforts, provided pro bono legal services in Portland, tenured with a U.S. District Court judge, and currently work with a law firm in Portland on injury, civil rights, and small business law. My law credentials for U.S. Senate include skills in negotiation and mediation for which I have been honored internationally in the U.S., Germany, and Canada. I am a member of the Oregon and California State Bars.

As a U.S. Senator, I would help restore self-determination to our local communities by ending intransigent partisanship and righting our capsized political system. The ideal democratic financial system must be revived in private and public sectors. The U.S. must return to a saner foreign policy, halting an international drain of American blood and treasure abroad while pursuing more effective strategies for peace. Perhaps most importantly, we must secure our sacred individual rights to life, liberty, and the pursuit of happiness – free of unwarranted government interference.

(This information furnished by Crawley for Oregon.)

# MEASURING CANDIDATE GENDER

- ▶ The new task:
  - ▶ Write code in Python to scrape the PDFs, pulling pictures out and parsing the text into a dataset
- ▶ Total time:
  - ▶ 3 months (not full time)
  - ▶ Time saved: .9 years
- ▶ Plus: now when future elections' PDFs come out, new data can be extracted right away

8 Candidates | Republican Candidates

United States Senator



**Mark Callahan**

Republican

**Occupation:** Information Technology (IT) Consultant  
**Occupational Background:** IT industry consulting/permanent positions, at local/national companies, over 13 years.

**Educational Background:** B.S. Business Administration / MIS, e-Law Minor, OSU – Corvallis, OR; MCP, MCSE Training Program, THE COMPUTER INSTITUTE – Rockville, MD  
**Prior Governmental Experience:** Marion County Republican Party PCP, 2013 – Present; Lane County Republican PCP, 2011-13; Eugene Income Tax Opposition Statement Committee, 11

**Leadership, Integrity, Character, Honor & Trust**

**Address Government Infringement:**

We must stop our government from violating our constitutional protections. From the NSA spying us, to themacking our 2nd Amendment rights, to the IRS targeting, to political correctness destroying our religious freedom. Our nation's sovereignty is under attack with the implementation of Agenda 21 and Common Core. As your constitutionally conservative Senator, I will stand strong, with Senator Ted Cruz for the rights of all Americans.

**conomy:**

As your Senator, who believes in the free market, I will work to make America's economy strong again, beginning with the Full Repeal of ObamaCare a.k.a "MerkleyCare".

**Balance Budget: Reduce Spending:**

As your fiscally conservative Senator, and a believer in limited government, I will reduce wasteful spending and push for a balanced budget amendment.

**Help Americans:**

We're at a crossroads affecting the very freedom our constitutional republic was founded upon. Supporting moderate candidates, that don't win general elections, is not an option. We need to elect a true conservative candidate that will fight for American values. We will win this election, by distancing ourselves from career politicians who are only looking to advance their own careers. We will win with a Pro-Life candidate that embraces our conservative values. A candidate not afraid to hold the President accountable for lying to us, without hesitating to convict him, if the House impeaches him. A candidate that isn't afraid to take on the establishment, promote freedom from government intrusion, dependency, poverty. My fellow Americans, I Am That Candidate. Thank you and May God Bless America.

Campaign website: [www.CallahanForOregon.com](http://www.CallahanForOregon.com)  
(This information furnished by Callahan For Oregon.)

United States Senator



**Timothy Crawley**

Republican

**Occupation:** Attorney  
**Occupational Background:** Nichols & Associate  
**Educational Background:** U.C. Hastings, J.D.; College, B.A.

**Prior Governmental Experience:** U.S. District Court, Oregon; Supreme Court of California; Federal Reserve

I was born and raised in Cottage Grove where my parents still reside. I consider it a personal obligation to help maintain its natural beauty and a healthy ecology by continuing to enjoy the benefits of a free market, without monopolistic power. I have travelled to nearly thirty countries and explored a multitude of foreign cultures and environments, but Oregon has always remained my home. That is why, after completing my legal education in California and beginning my professional career there, I have returned to Oregon.

My family always encouraged me to serve others. I advocated for civil rights in the Middle East; crewed on an Alaskan fishing boat; educated the public for the Federal Reserve Bank; and represented tech entrepreneurs in Silicon Valley. Since my return to Oregon, I have been a volunteer in environmental restoration efforts, provided pro bono legal services in Portland, tenured with a U.S. District Court judge, and currently work with a law firm in Portland on injury, civil rights, and small business law. My law credentials for U.S. Senate include skills in negotiation and mediation for which I have been honored internationally in the U.S., Germany, and Canada. I am a member of the Oregon and California State Bars.

As a U.S. Senator, I would help restore self-determination in our local communities by ending intransigent partisanship and righting our capsized political system. The ideal democratic financial system must be revived in private and public sectors. The U.S. must return to a saner foreign policy, halting an international drain of American blood and treasure abroad while pursuing more effective strategies for peace. Perhaps most importantly, we must secure our sacred individual rights to life, liberty, and the pursuit of happiness – free of unwarranted government interference.

(This information furnished by Crawley for Oregon)

## PRACTICAL EXAMPLES

---

# WORKING WITH RAs

### ▶ The task:

- ▶ RAs had to find specific photos of presidential candidates using different search terms
- ▶ Different RAs used different labeling schemes, and pictures for all candidates were saved into one big folder
- ▶ Needed consistent labeling scheme and separate folders

### ▶ Time to fix by hand:

- ▶ 200 candidate photos x
- ▶ 5 presidential candidates x
- ▶ 15 seconds a photo to re-label and move =
- ▶ 4.2 hours fixing

Date	File Name	Date Modified
Apr 15, 2016	2016 president primary 2.jpg	Apr 15, 2016
Apr 22, 2016	clinton 2016 5.jpg	Apr 22, 2016
Apr 22, 2016	clinton democrat 1.jpg	Apr 22, 2016
Apr 22, 2016	clinton democrat 7.jpg	Apr 22, 2016
Apr 22, 2016	clinton msnbc 6.jpg	Apr 22, 2016
Apr 22, 2016	clinton pbs 3.jpg	Apr 22, 2016
Apr 22, 2016	clinton pbs 5.jpg	Apr 22, 2016
Apr 22, 2016	clinton pbs 7.jpg	Apr 22, 2016
Apr 22, 2016	clinton president 2.jpg	Apr 22, 2016
Apr 22, 2016	clinton president 7.jpg	Apr 22, 2016
Apr 22, 2016	clinton primary 10.jpg	Apr 22, 2016
Apr 22, 2016	clinton wpost 2.jpg	Apr 22, 2016
Apr 22, 2016	clinton wpost 7.jpg	Apr 22, 2016
Apr 18, 2016	co bernie sanders 2016 7.jpg	Apr 18, 2016
Apr 18, 2016	co bernie sanders 2016 12.jpeg	Apr 18, 2016
Apr 18, 2016	co bernie sanders 2016 13.png	Apr 18, 2016
Apr 18, 2016	co bernie sanders 2016 17.jpeg	Apr 18, 2016
Apr 17, 2016	co bernie sanders abc 12.jpg	Apr 17, 2016
Apr 18, 2016	co bernie sanders democrat 5.jpeg	Apr 18, 2016
Apr 15, 2016	co bernie sanders fox news 6.png	Apr 15, 2016
Apr 17, 2016	co bernie sanders pbs 1.jpeg	Apr 17, 2016
Apr 17, 2016	co bernie sanders pbs 16.jpg	Apr 17, 2016
Apr 18, 2016	co bernie sanders president 4.png	Apr 18, 2016
Apr 18, 2016	co bernie sanders president 5.jpeg	Apr 18, 2016
Apr 18, 2016	co bernie sanders primary 5.jpeg	Apr 18, 2016

# WORKING WITH RAS

- ▶ **The new task:**

- ▶ Create consistent labeling scheme for filenames and move into separate folders using Bash

- ▶ **Total time:**

- ▶ **1.25 hours coding**
- ▶ 3 hours saved (again, not accounting for human error or fatigue)

e	Date Mo
clinton usatoday 2.jpg	Apr 22,
clinton democrat 11.jpg	Apr 22,
clinton president 4.jpg	Apr 22,
clinton 2016 3.jpg	Apr 22,
clinton 2016 10.jpg	Apr 22,
clinton president 3.jpg	Apr 22,
clinton 2016 2.jpg	Apr 22,
clinton republican 6.jpg	Apr 22,
clinton nytimes 4.jpg	Apr 22,
clinton republican 3.jpg	Apr 22,
clinton wpost 6.jpg	Apr 22,
clinton republican 7.jpg	Apr 22,
clinton wpost 10.jpg	Apr 22,
clinton democrat 2.jpg	Apr 22,
clinton wpost 1.jpg	Apr 22,
clinton primary 4.jpg	Apr 22,
clinton democrat 6.jpg	Apr 22,
clinton wpost 13.jpg	Apr 22,
clinton wpost 11.jpg	Apr 22,
clinton president 8.jpg	Apr 22,
clinton wpost 3.jpg	Apr 22,
clinton wpost 5.jpg	Apr 22,
clinton abc 2.jpg	Apr 22,
clinton nbc 10.jpg	Apr 22,
clinton primary 8.jpg	Apr 22,

# FINAL PROJECTS

- ▶ Previous final projects include:
  - ▶ Using Python and R to collect Facebook API data on reactions to political candidates, and R to analyze and visualize the data
  - ▶ Using Python to scrape news articles about 9/11, and R to perform text analysis on them and generate visualizations
  - ▶ Using APIs in R to post a survey about local elections to MTurk, and analyzing and plotting the results in R
- ▶ Your mileage will vary depending on how much coding experience you already have and how advanced your research agenda is

# OBSTACLES TO LEARNING

- ▶ Unspoken/tacit knowledge among coders ("banana banana banana")
- ▶ Learning to think like a computer (1s and 0s not as easy as you think)
- ▶ Becoming impatient while learning basic concepts (most errors lie here)
- ▶ Low endurance (the diligent bird builds the worm)
- ▶ Infrequent practice (fluency doesn't come from weekends)
- ▶ Solo coding (teaching is learning)
- ▶ Copy-and-pasting instead of typing (ctrl+c is not your friend)
- ▶ Disorganized coding (would you write an essay without an outline?)
- ▶ Failing to comment code as you go (## what does this do?)
- ▶ Expecting to be Gandalf (10-20 lines a day)

### CLASS =

- ▶ Participation (25%)
  - ▶ Be kind to your classmates. We're all learning. Even me. (Especially me.)
- ▶ Homework (50%)
  - ▶ Every week. Don't be late. (You wouldn't like me when I'm angry.)
- ▶ Final project (25%)
  - ▶ Feasibility is your friend. Feb. 28 proposal, April presentations.

# LOGISTICS

### ▶ **Lecture and section**

- ▶ “Workshop style:” you will be programming every class, so bring your laptop
- ▶ You will often be pair-programming on in-class challenges
- ▶ Section is **\*required\***; it focuses on review as needed, but is primarily time to work on homework in groups (and to present final projects in April)
- ▶ Jae Yeon Kim is technical assistant to the course

### ▶ **Next assignments**

- ▶ All assignments will be turned in via bCourses
- ▶ Assignment 1 - get software installed (instructions on bCourses/github)
- ▶ Assignment 2 - learn command line on codeacademy
- ▶ Longer assignments worth more points (see bCourses), so start those early

## THANKS

---

SPECIAL THANKS FOR COURSE MATERIALS ARE  
DUE TO **ROCHELLE TERMAN**, THE CREATOR OF  
PS239T AND INNOVATRIX EXTRAORDINAIRE.