Today (10/19/20)

• Op-Ed due by **MIDNIGHT ON WEDNESDAY**
• No reading for next class (Todd Rogers guest lectures – read his paper if you haven’t already)
• Note new syllabus ordering!
• Today:
  – Lecture – Data and elections 3
  – Student Presentations
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Speaker</th>
<th>Date</th>
<th>Topic</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-31</td>
<td>Introduction</td>
<td>Fran</td>
<td>9-3</td>
<td>The Data-driven world</td>
<td>Fran</td>
</tr>
<tr>
<td>9-10</td>
<td>Data and COVID-19 - models</td>
<td>Fran</td>
<td>9-14</td>
<td>Data and COVID-19 – contact tracing</td>
<td>Fran</td>
</tr>
<tr>
<td>9-17</td>
<td>Data and the Opioid Crisis</td>
<td>Liz Chiarello</td>
<td>9-21</td>
<td>Data and Privacy - Intro</td>
<td>Fran</td>
</tr>
<tr>
<td>9-24</td>
<td>Data and Privacy – Differential Privacy and the Census</td>
<td>Fran</td>
<td>9-28</td>
<td>Data and Privacy – Anonymity</td>
<td>Fran</td>
</tr>
<tr>
<td>10-1</td>
<td>Data and Privacy - Law</td>
<td>Fran</td>
<td>10-5</td>
<td>Digital rights in the EU and China</td>
<td>Fran</td>
</tr>
<tr>
<td>10-8</td>
<td>Data and Elections 1</td>
<td>Fran</td>
<td>10-12</td>
<td>NO CLASS – Columbus / Indigenous Peoples’ Day</td>
<td>Fran</td>
</tr>
<tr>
<td>10-15</td>
<td>Data and Elections 2</td>
<td>Fran</td>
<td>10-19</td>
<td>Data and Elections 3</td>
<td>Fran</td>
</tr>
<tr>
<td>10-22</td>
<td>Data and Elections 4</td>
<td>Todd Rogers</td>
<td>10-26</td>
<td>Data and Research 1</td>
<td>Fran</td>
</tr>
<tr>
<td>10-29</td>
<td>Data and Research 2</td>
<td>Josh Greenberg</td>
<td>11-2</td>
<td>Data and Discrimination 1</td>
<td>Fran</td>
</tr>
<tr>
<td>11-5</td>
<td>Data and Discrimination 2</td>
<td>Fran</td>
<td>11-9</td>
<td>Data and the IoT 1</td>
<td>Fran</td>
</tr>
<tr>
<td>11-12</td>
<td>Data and the IoT 2</td>
<td>Fran</td>
<td>11-15</td>
<td>Data and Ethics</td>
<td>Fran</td>
</tr>
<tr>
<td>11-23</td>
<td>Cybersecurity</td>
<td>Bruce Schneier</td>
<td>12-3</td>
<td>Data Science</td>
<td>Fran</td>
</tr>
<tr>
<td>11-30</td>
<td>Data Infrastructure</td>
<td>Fran</td>
<td>12-7</td>
<td>Data Careers</td>
<td>Kathy Pham</td>
</tr>
<tr>
<td>12-3</td>
<td>Data Science</td>
<td>Fran</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12-10</td>
<td>Wrap-up</td>
<td>Fran</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lecture and Discussion

• Election Models in 2016
• The 2020 Election
Data and the 2016 Election – Failure of Big Data?

Who will win the presidency?

Hillary Clinton: 71.4%
Donald Trump: 28.6%

Electoral votes
- Hillary Clinton: 302.2
- Donald Trump: 235.0
- Evan McMullin: 6.5
- Gary Johnson: 6.5

Popular vote
- Hillary Clinton: 48.5%
- Donald Trump: 44.4%
- Gary Johnson: 5.0%
- Other: 2.5%
What Happened?
Why were most predictions inaccurate?

- Were the models wrong?
- Were the interpretations wrong?
- Was the data faulty?
- Was the sampling / polls wrong?
- Was voter behavior just one of the low probability outcomes?

Map from http://www.270towin.com/2016_Election/interactive_map
### How people voted: Exit Polls and Election Results

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Breakdown(s)</th>
</tr>
</thead>
</table>
| **Age**                 | • 18-29: 55% for Clinton, 37% for Trump  
• 30-44: 50% for Clinton, 42% for Trump  
• 45+: 53% for Trump  
• 45-64: 44% for Clinton  
• 65+: 45% for Clinton                       |
| **Gender**              | • 54% of women voted for Clinton; 42% of women voted for Trump  
• 53% of men voted for Trump; 41% of men voted for Clinton                                       |
| **Ethnicity**           | • White voters: 58% for Trump, 37% for Clinton  
• Black voters: 88% for Clinton; 8% for Trump  
• Hispanic and Asian voters: 65% for Clinton; 29% for Trump                                           |
| **Education**           | • College grads: 49% for Clinton  
• Postgrads: 58% for Clinton  
• High school or less: 51% for Trump  
• Some college / Associate degree: 52% for Trump                                                |
| **Religion**            | • Catholic: 52% for Trump  
• Protestants / Christians: 58% for Trump  
• Jewish: 71% for Clinton  
• Other: 62% for Clinton  
• No religion: 68% for Clinton                                                               |
| **Income**              | • Under $30K: 53% for Clinton  
• $30K-$49.99K: 51% for Clinton  
• $50K - $99.99K: 50% for Trump  
• $100K-$199.99K: 48% for Trump  
• $250K+: 48% for Trump                                                                         |
| **Locale (Urban vs. Rural)** | • Cities with > 50K residents: 59% for Clinton, 35% for Trump  
• Rural areas: 62% for Trump, 34% for Clinton  
• Suburbs: 50% for Trump, 45% for Clinton                                                      |
Who can / did vote?

All Americans
320,000,000+

Voting age population
251,107,404 (78.5%)

Eligible voters
231,556,622 (72.4%)

Registered voters
~200,000,000 (62.5%)

Voters
138,884,643
(43.4% of all citizens, 60% of eligible voters)

Statistics from:
Election models: Model and Interpretation
Accuracy

Many challenges in modeling and interpretation:

- Raw polling data supplemented by estimates on how many people will vote and what undecided voters will do.

- Historical inferences about past patterns of turnout, demographics, economic conditions and party loyalty may not be accurate for present day.

- If polls shows that candidate “wins” by a small margin within the margin of error, it is risky to interpret this as a “win.”

From: https://projects.fivethirtyeight.com/2016-election-forecast/

Data Integrity – Was poll data accurate?

• Many suspected that people lied about voting for Trump

• Trafalgar Group’s approach to improving data accuracy -- Adjust numbers to account for people’s hesitance to admit a Trump vote
  – Used robotic calls for which Trump voters seemed more comfortable
  – Added a “neighbor” question -- Who do you think your neighbors will vote for? – and checked to see if the numbers were different
  – Created a demographic of people who had not voted in 6+ years but planned to vote for Trump

• Trafalgar predicted Trump win in Pennsylvania and Michigan (but not all states)
Sampling Accuracy

Key sampling questions

• How representative is the sample of population?
• How big is the sample / what is the margin of error?
• How biased are the sampling vehicles – land lines, human interviews, tweets, etc.?
• How representative is the sample of turnout? For eligible voters? For eligible voters who actually vote?
• How accurate is the data (are people lying)?

Figure from http://www.forbes.com/sites/startswithabang/2016/11/09/the-science-of-error-how-polling-botched-the-2016-election/#75748a437da8

A visualization of how your statistical uncertainty drops as your sample size increases. Image credit: Foolproof at English Wikipedia.
Who made correct predictions?

- **Investor’s Business Daily (IBD/TIPP poll)**
  
  **Predicted:** Trump would win by 1.6%
  
  **Approach:**
  
  - Start with random sample from public, adjust for census statistics and age, gender, religion, look at registered and likely voters, adjust for party registrations, **enthusiasm**
  - Poll made more calls to **smartphones** than landlines.
  - People represented wide range of people in the country (including a representative sample of types of phones used)
  - Poll questioned respondents about enthusiasm and factored this into results

- **USC / LA Times poll (USC economics prof Arie Kapteyn)**

  **Predicted:** Trump would win by 3%
  
  **Approach:**
  
  - Pollsters sought to **balance both big groups** (e.g. men and women) and **smaller groups** (e.g. young minority voters)
  - **Weighting of responses** in polls used to make them more fully representative. Sample includes representation of demographic statistics including race, gender, age.
Who made correct predictions?

- Alan Lichtman / American University historian

Predicted: Trump wins

Approach:

- Developed 13 T/F keys that predict election outcome. True favors incumbent party. If 6+ are false, change is predicted.
- Has worked in every election for the last 30 years.

- 2020: Predicts Trump will lose

- Lichtman’s Keys:

  1. **Party Mandate**: After the midterm elections, the incumbent party holds more seats in the U.S. House of Representatives than after the previous midterm elections.
  2. **Contest**: There is no serious contest for the incumbent party nomination.
  3. **Incumbency**: The incumbent party candidate is the sitting president.
  4. **Third party**: There is no significant third party or independent campaign.
  5. **Short-term economy**: The economy is not in recession during the election campaign.
  6. **Long-term economy**: Real per capita economic growth during the term equals or exceeds mean growth during the previous two terms.
  7. **Policy change**: The incumbent administration effects major changes in national policy.
  8. **Social unrest**: There is no sustained social unrest during the term.
  9. **Scandal**: The incumbent administration is untainted by major scandal.
  10. **Foreign/military failure**: The incumbent administration suffers no major failure in foreign or military affairs.
  11. **Foreign/military success**: The incumbent administration achieves a major success in foreign or military affairs.
  12. **Incumbent charisma**: The incumbent party candidate is charismatic or a national hero.
  13. **Challenger charisma**: The challenging party candidate is not charismatic or a national hero.
Election 2020: Some predictions

- **The Primary Model**, Helmut Norpoth, Stonybrook (http://primarymodel.com/)
- **2016 Prediction**: Trump wins
- **2020 Prediction**: Trump wins
- **Approach**:
  - Statistical model that uses primaries rather than polls to predict outcomes.
  - Takes “swing of the electoral pendulum” into consideration (Republicans favored after two democratic terms)
- **Primary model has picked the winner of 25 out of 27 presidential elections since presidential primaries were introduced**
Election 2020: Some predictions

- **The Economist Model**
  (https://projects.economist.com/us-2020-forecast/president)
- **Prediction:** Biden wins
- **Approach:**
  - Model focuses on variety of national polls and economic indicators.
  - Parameter sweep comparisons of 20,000 data variations accounting for
    - Polling error
    - Changes in turnout / political environment / effects of campaigning
  - You can download models and data!
Election 2020 Discussion

Lecture 13 Sources 1


- “The trouble is not with polling but with the limits to human interpretation of data,” Quartz, http://qz.com/832908/confirmation-bias-is-why-we-couldnt-predict-a-trump-victory/

- “There are Many Ways to Map election Results. We’ve Tried Most of Them.”, NY Times, http://www.nytimes.com/interactive/2016/11/01/upshot/many-ways-to-map-election-results.html?_r=0


• Presentations for October 22

• Presentations for October 26

• Presentations for October 29
Need Volunteers – Presentations for November 2


Presentations for Today


• “Some states have embraced on-line voting. It’s a huge risk.”, Politico, https://www.politico.com/news/2020/06/08/online-voting-304013 (Xilin)