Gov. 1010 Mini-Lecture

Why is it important to think about how data is created?: A Real-world Example

Mark Hill[1]

[1]Department of Government
Harvard University

October 30, 2019

Outline

- Why present this data in this class?

Outline

- Why present this data in this class?
- Motivate thinking about the data you are given

Outline

- Why present this data in this class?
- Motivate thinking about the data you are given
- Present real-world data that made some headlines

Outline

- Why present this data in this class?
- Motivate thinking about the data you are given
- Present real-world data that made some headlines
- Explore the data a bit

Outline

- Why present this data in this class?
- Motivate thinking about the data you are given
- Present real-world data that made some headlines
- Explore the data a bit
- Ask you to tell me what you think

Outline

- Why present this data in this class?
- Motivate thinking about the data you are given
- Present real-world data that made some headlines
- Explore the data a bit
- Ask you to tell me what you think
- A deeper look at the data and an explanation

Why present this data?

- Data from a real-world survey that provides a lot of interesting material

Why present this data?

- Data from a real-world survey that provides a lot of interesting material
- Data science/analytics aspects for those interested

Why present this data?

- Data from a real-world survey that provides a lot of interesting material
- Data science/analytics aspects for those interested
- Fascinating political takeaways

Why present this data?

- Data from a real-world survey that provides a lot of interesting material
- Data science/analytics aspects for those interested
- Fascinating political takeaways
- Something for you to consider when designing your own surveys

Motivation

Given a dataset, most people assume the data is fine and they dive right into their analysis without thinking.

Motivation

Given a dataset, most people assume the data is fine and they dive right into their analysis without thinking.

- How was the data collected?

Motivation

Given a dataset, most people assume the data is fine and they dive right into their analysis without thinking.

- How was the data collected?
- What individuals or observations are in the data?

Motivation

Given a dataset, most people assume the data is fine and they dive right into their analysis without thinking.

- How was the data collected?
- What individuals or observations are in the data?
- Where is the data from?

Motivation

Given a dataset, most people assume the data is fine and they dive right
into their analysis without thinking.

- How was the data collected?
- What individuals or observations are in the data?
- Where is the data from?
- Do I trust this data and, more importantly, will my audience be
  convinced if I use this data to support my claims?

Motivation

Given a dataset, most people assume the data is fine and they dive right into their analysis without thinking.

- How was the data collected?
- What individuals or observations are in the data?
- Where is the data from?
- Do I trust this data and, more importantly, will my audience be convinced if I use this data to support my claims?

It's often very useful to think through these questions before you begin any type of statistical analysis.

Real-world Data

- The Harris Poll carried out by Harris Insights and Analytics in collaboration with the Harvard Center for American Political Studies.
- The poll was a self-administered online questionnaire of 2,148 registered voters conducted between February 11-13, 2017.
- Respondents were asked a variety of questions about their approval of different political figures and policies, along with demographic questions.

Real-world Data

- The Harris Poll carried out by Harris Insights and Analytics in collaboration with the Harvard Center for American Political Studies.
- The poll was a self-administered online questionnaire of 2,148 registered voters conducted between February 11-13, 2017.
- Respondents were asked a variety of questions about their approval of different political figures and policies, along with demographic questions.

Harris Insights and Analytics conducts the poll based on online methodologies they have been using for more than a decade and the results are donated to the Harvard Center for American Political Studies.

The results are represented as the results of the questions asked according to The Harris Poll methodology.

Published Results

Poll: Americans overwhelmingly oppose sanctuary cities

BY JONATHAN EASLEY - 02/21/17 02:19 PM EST

3,007 COMMENTS

71,256 SHARES

Just In...

"According to one recent poll, 80 percent of Americans believe that cities that arrest illegal immigrants for a crime should be required to turn them over to immigration authorities."

— **Attorney General Jeff Sessions, news briefing, March 27, 2017**

"I think the last poll I saw on this issue, on sanctuary cities, was somewhere in the 80 percent that American people don't support sanctuary cities, they don't want their tax dollars used to finance people who are in this country illegally."

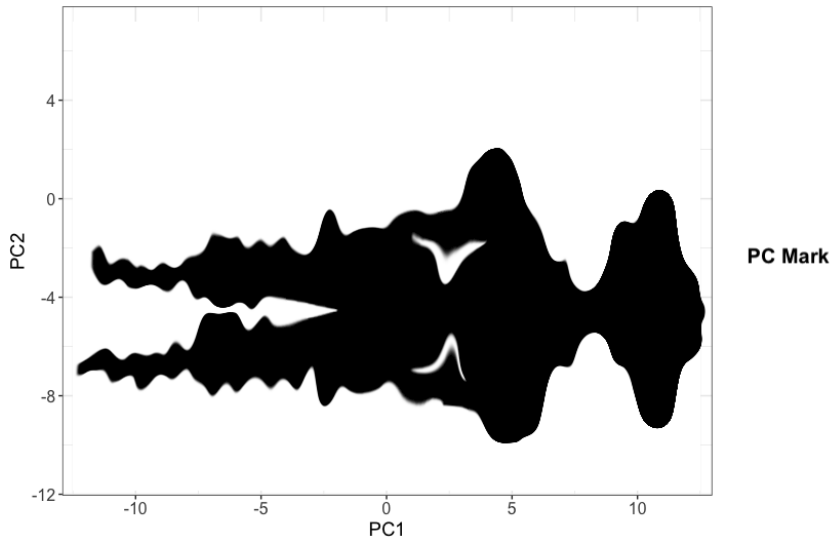— **White House press secretary Sean Spicer, news briefing, March 14**

An Intuitive Explanation of Principle Component Analysis
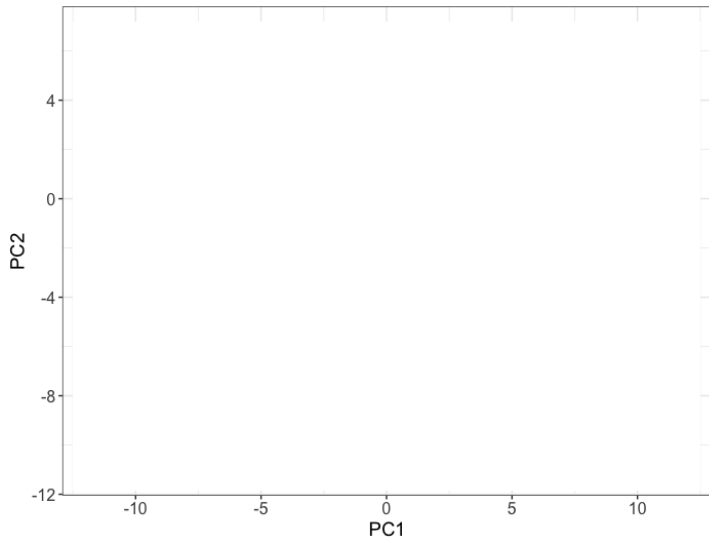
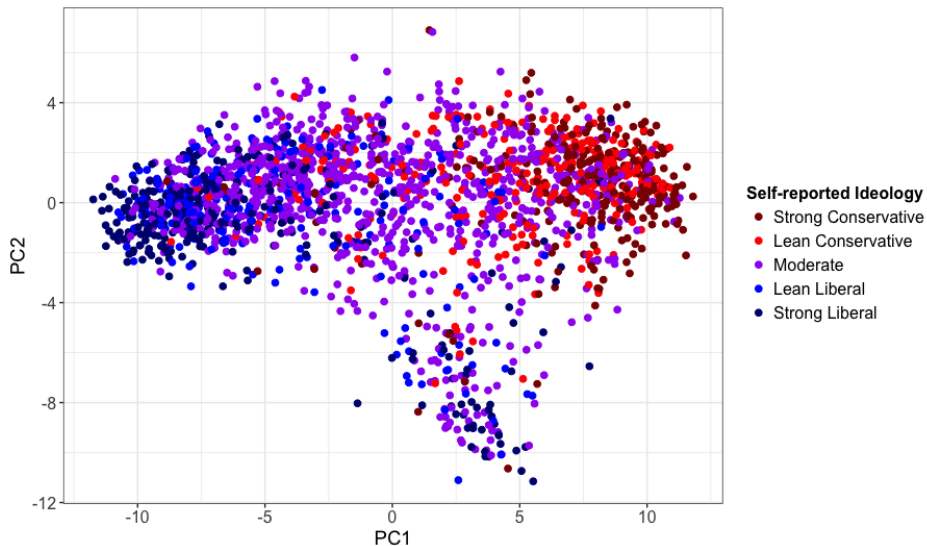# An Intuitive Explanation of Principle Component Analysis
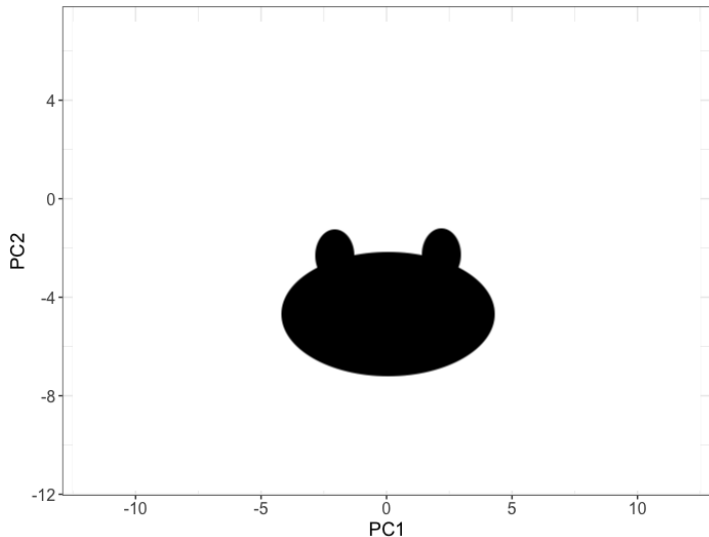
An Intuitive Explanation of Principle Component Analysis
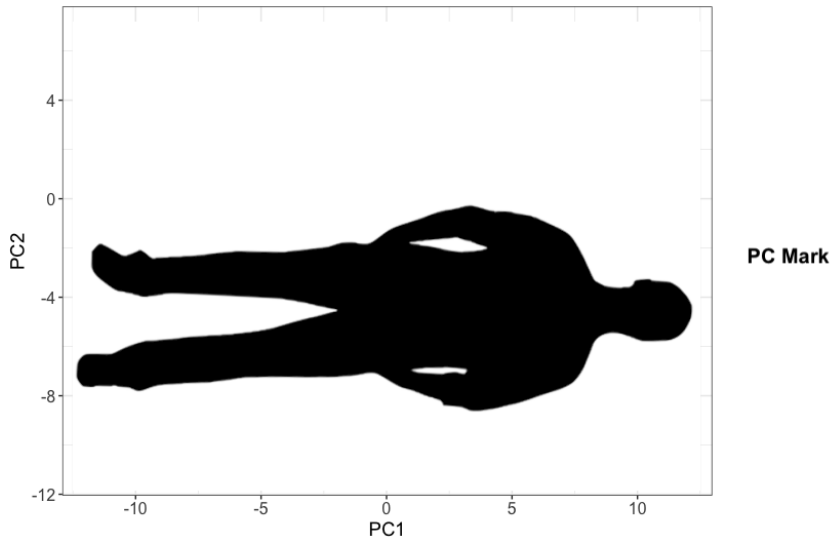
An Intuitive Explanation of Principle Component Analysis

Data Exploration: PCA

Data Exploration: Most Influential Variables in PC 1 and 2

For PC1:

- Trump job approval
- Trump favorable/unfavorable view
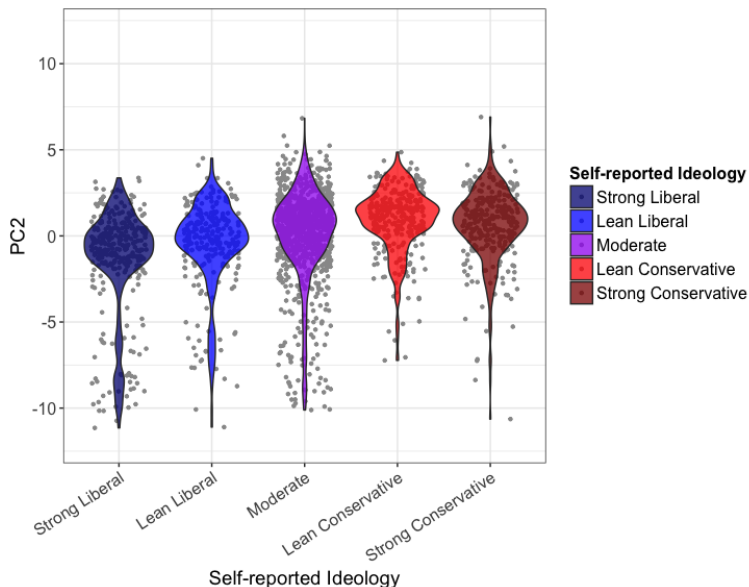- Trump promise to stop immigration from Syria
- Obamacare
- Building the wall

For PC2:

- Strength of US Economy
- How is US perceived in eyes of the world
- Progress in reducing budget deficit
- Affirmative action
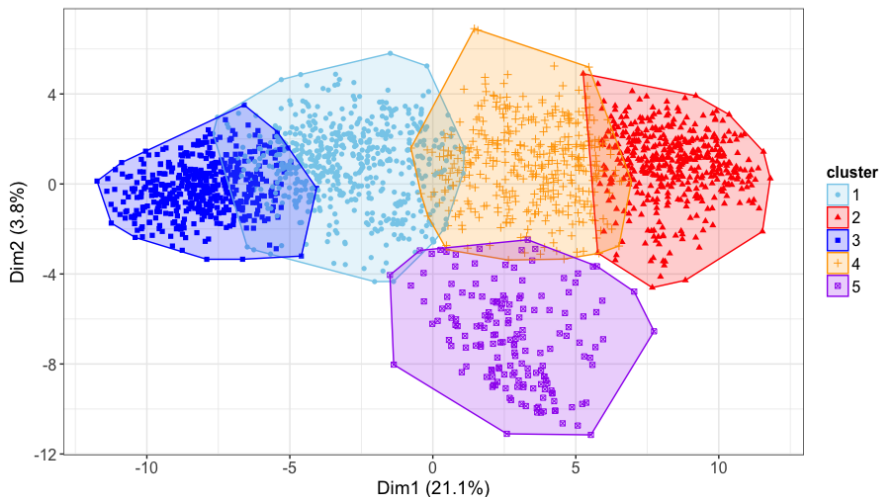- Another question on affirmative action

Data Exploration: PC1
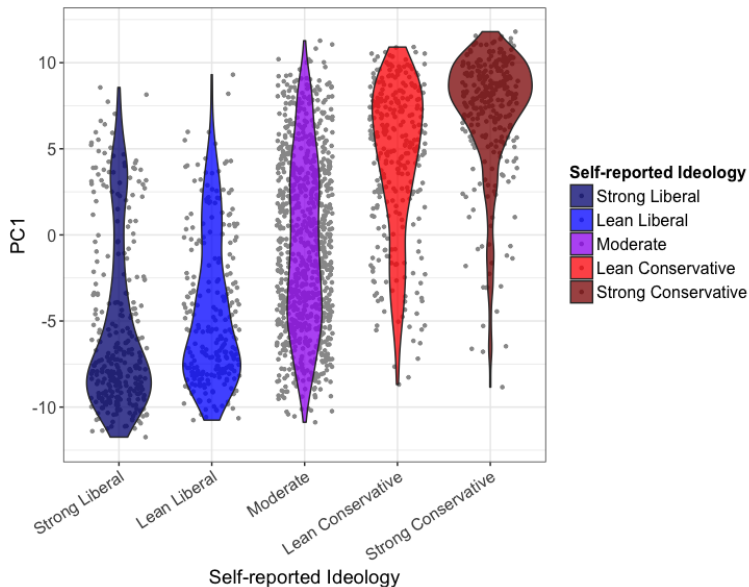
Data Exploration: PC2

Clustering - Let the computer assign groups

About 8% of the data is in the purple cluster (164 obs.)

Data Exploration: Demographics

Cluster	Blue (Liberals)	Sky Blue (Lean Lib.)	Orange (Lean Cons.)	Red (Conservative)	Purple (??)	All
Ideology (5-pt scale) 5=str. conservative	1.97	2.79	3.38	4.16	2.45	2.99
Party Identification (%)						
Republican	1.2	8.7	44.3	63.9	40.9	28.4
Democratic	71.4	53.4	16.0	4.0	44.5	38.9
Independent	24.9	30.7	30.5	28.2	14.0	27.4
Vote (%)						
Trump	0.2	2.1	51.4	88.5	57.9	34.5
Clinton	88.4	60.2	8.4	0.2	26.2	41.1
Other	4.4	10.6	12.5	2.7	4.3	7.2
Refused	7.1	27.0	27.7	8.6	11.6	17.2
N	482	566	393	451	164	2056

What's going on?

What's going on?

- Discontented Democrats?

- Discontented Democrats?
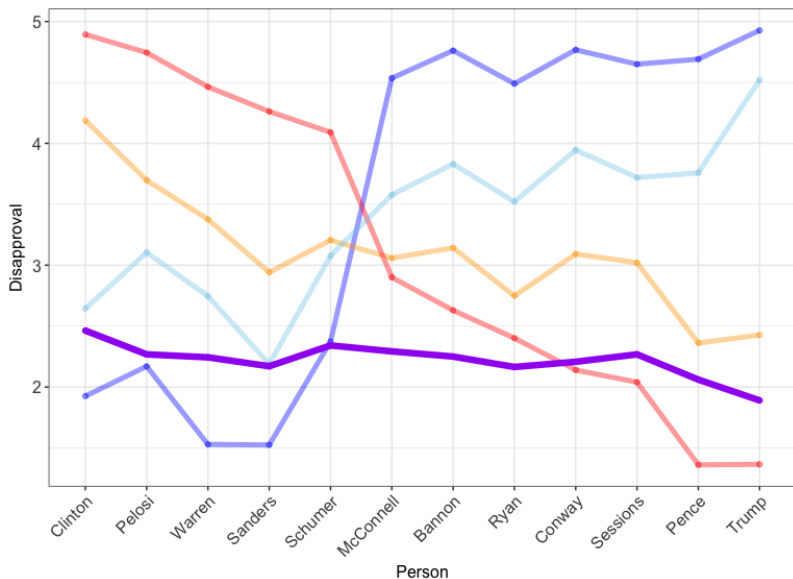- Socially or fiscally conservative, but liberal in other aspects?

What's going on?

- Discontented Democrats?
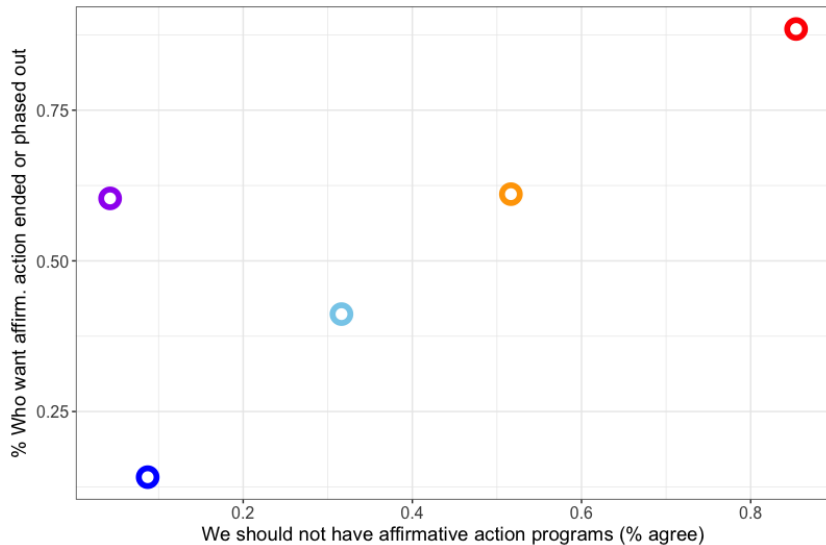- Socially or fiscally conservative, but liberal in other aspects?
- Single-issue voters?

- Discontented Democrats?
- Socially or fiscally conservative, but liberal in other aspects?
- Single-issue voters?
- Something else?

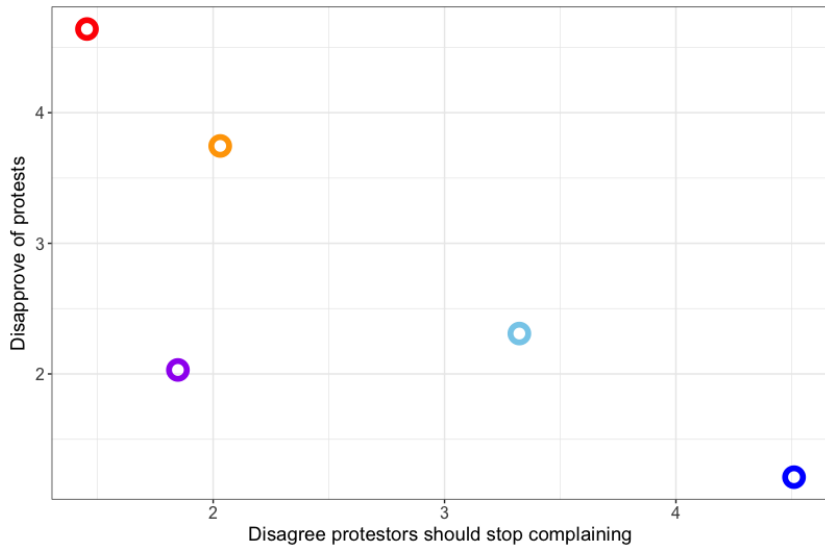# A Deeper Look: Approval

A Deeper Look: Affirmative Action

A Deeper Look: Protests

How did this happen?

How did this happen?

The ordering of response choices for most questions was not randomized!

How did this happen?

The ordering of response choices for most questions was not randomized!

I41. Which comes closer to your point of view regarding affirmative action programs in the work place?
1. We should have affirmative action programs to overcome past discrimination.
2. We should have affirmative action programs to increase diversity.
3. We should not have affirmative action programs.

I42. What do you think should happen to affirmative action programs? Should they be ended now, should they be phased out over the next few years, or should affirmative action programs be continued for the foreseeable future?
1. Ended Now
2. Phased Out
3. Continued
4. Unsure

Does this matter?