SMI 606: Introduction to Quantitative Data Analysis

Dr. Todd K. Hartman
Sheffield Methods Institute

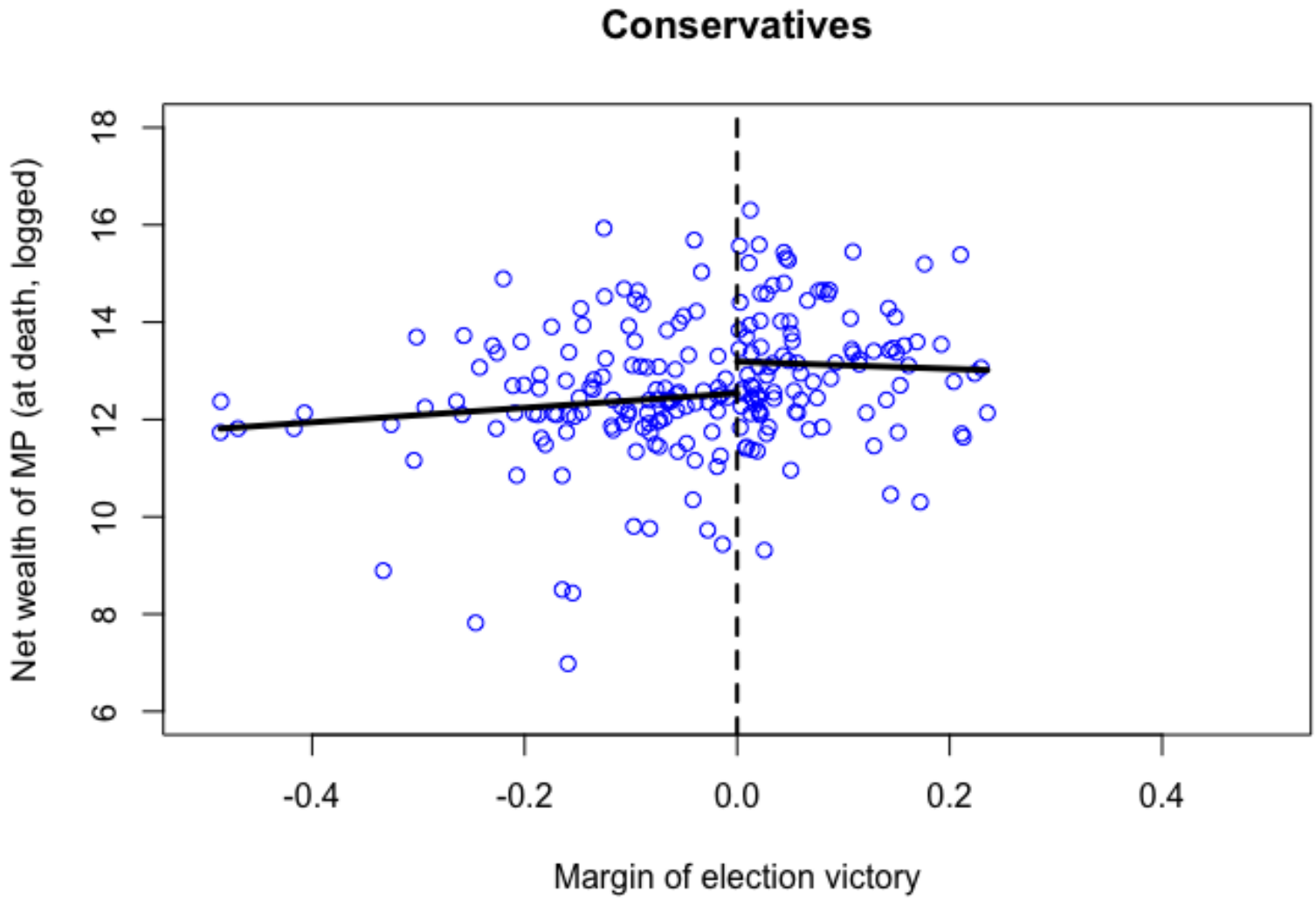
Welcome!

Keep Calm

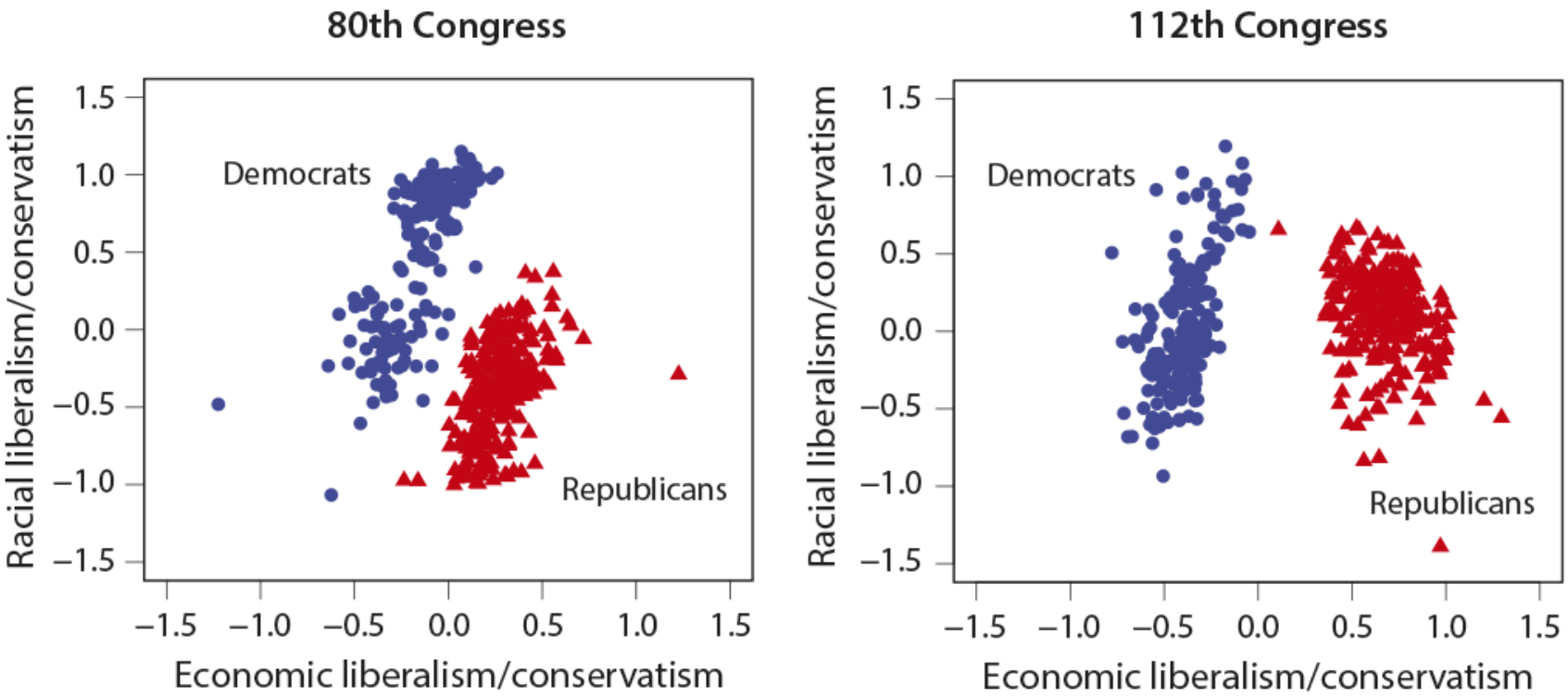
Roadmap: Week 1

- Some cool things you'll learn!
- Module overview
- Lecture: How do we use data?
- Lab: Getting started in R

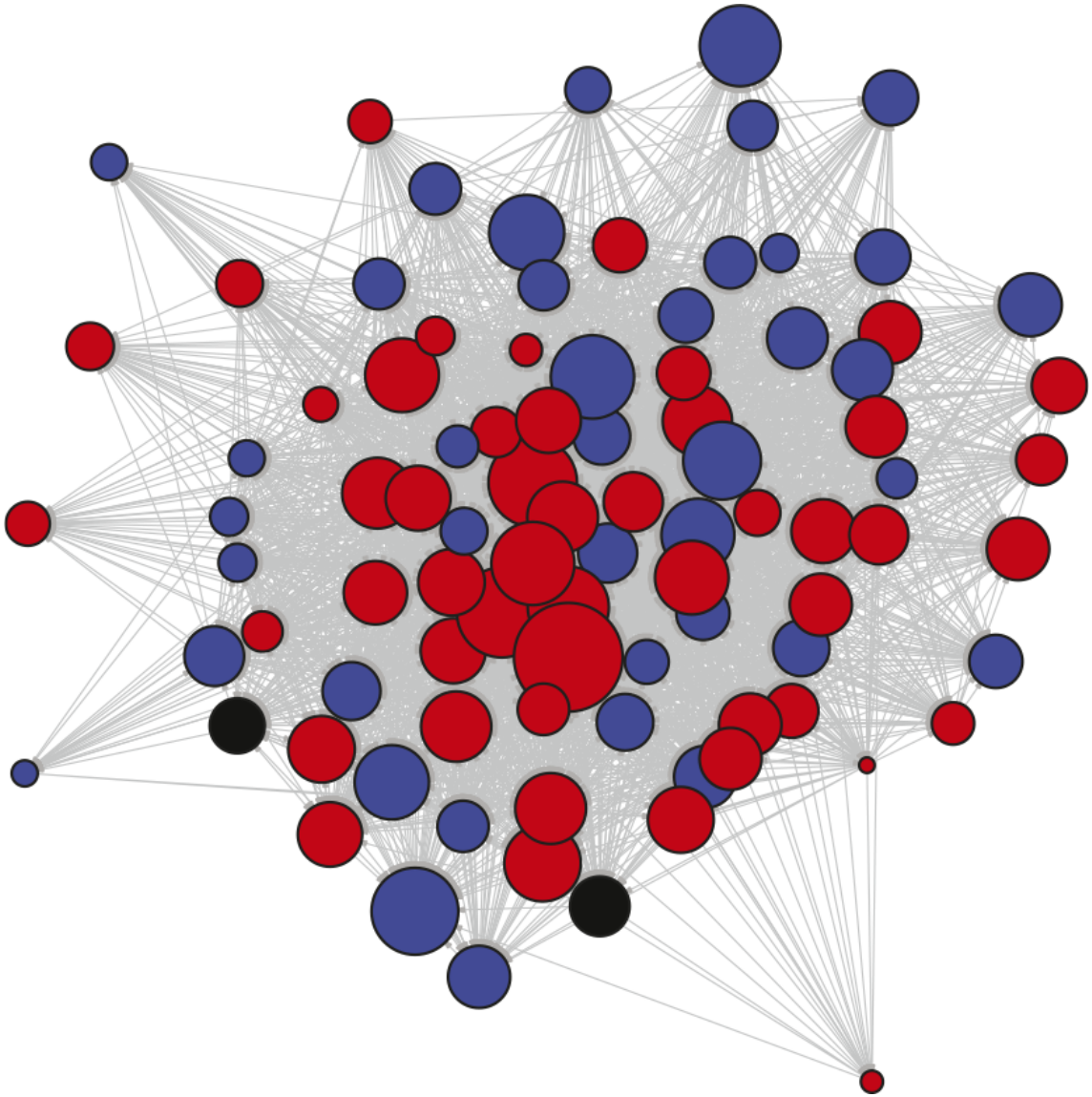
Regression


Conservatives
Net wealth of MP (at death, logged)
6
8
10
12
14
16
18
-0.4
-0.2
0.0
0.2
0.4
Margin of election victory

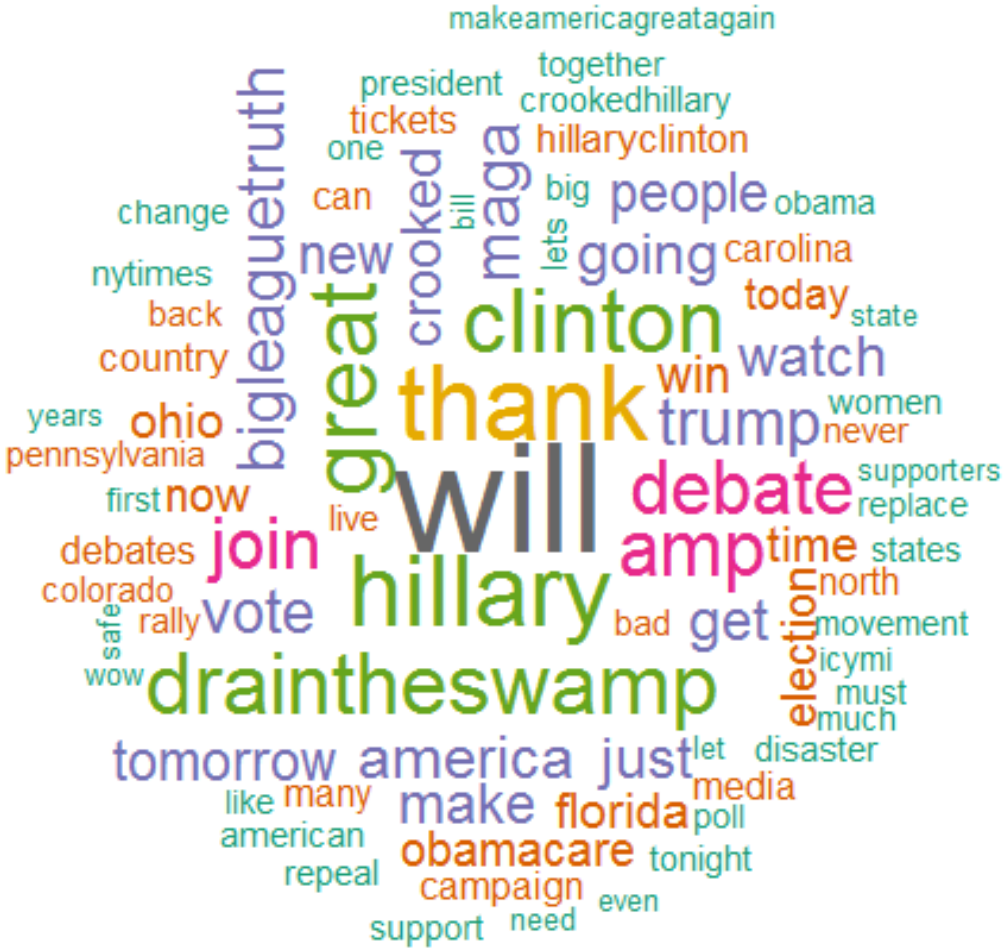
Cluster Analysis



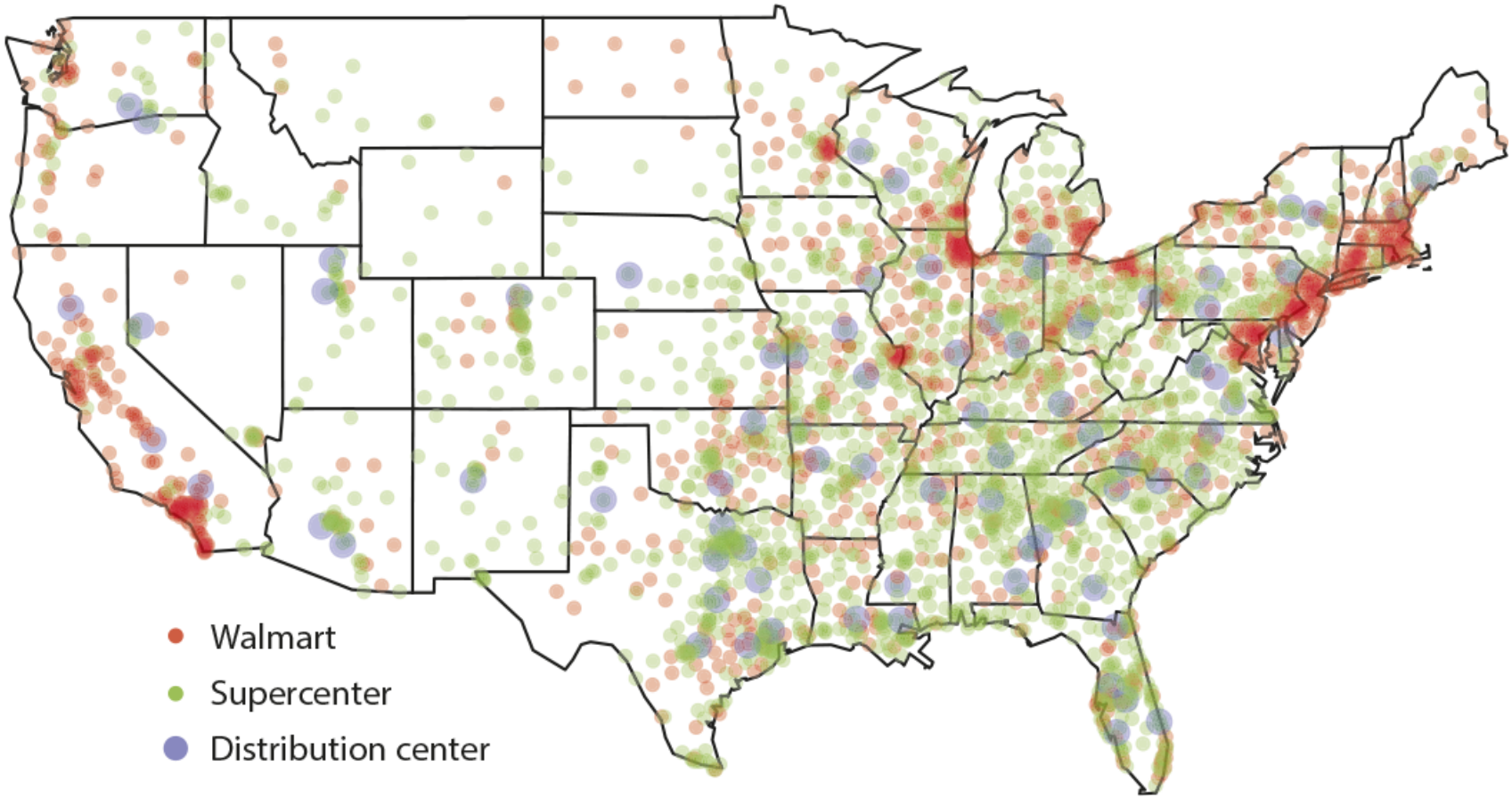
Social Network Analysis



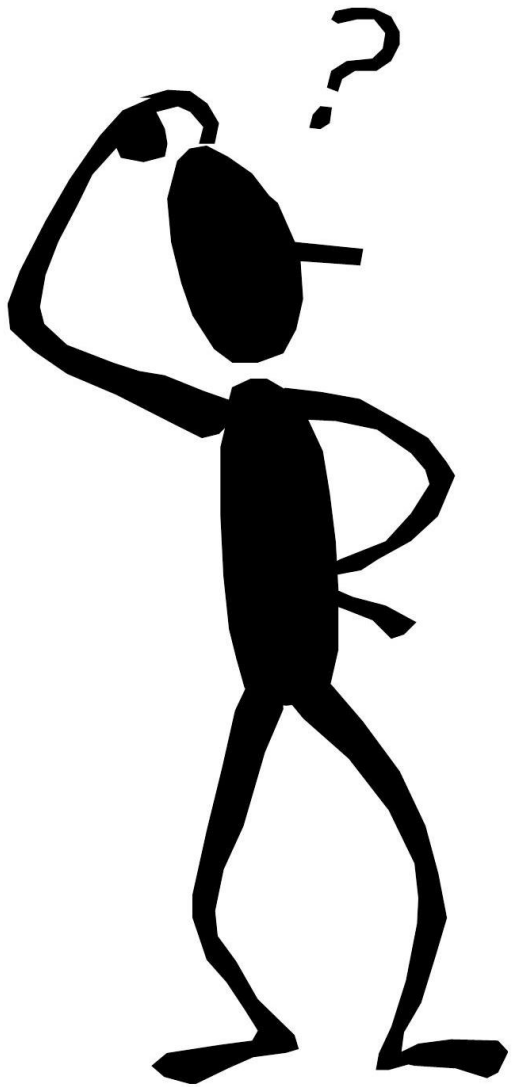
Text Analysis



Spatial Analysis



How Do We Use Data?



How Do We Use Data?

1. Description
    - e.g., census demographics, election results, etc.

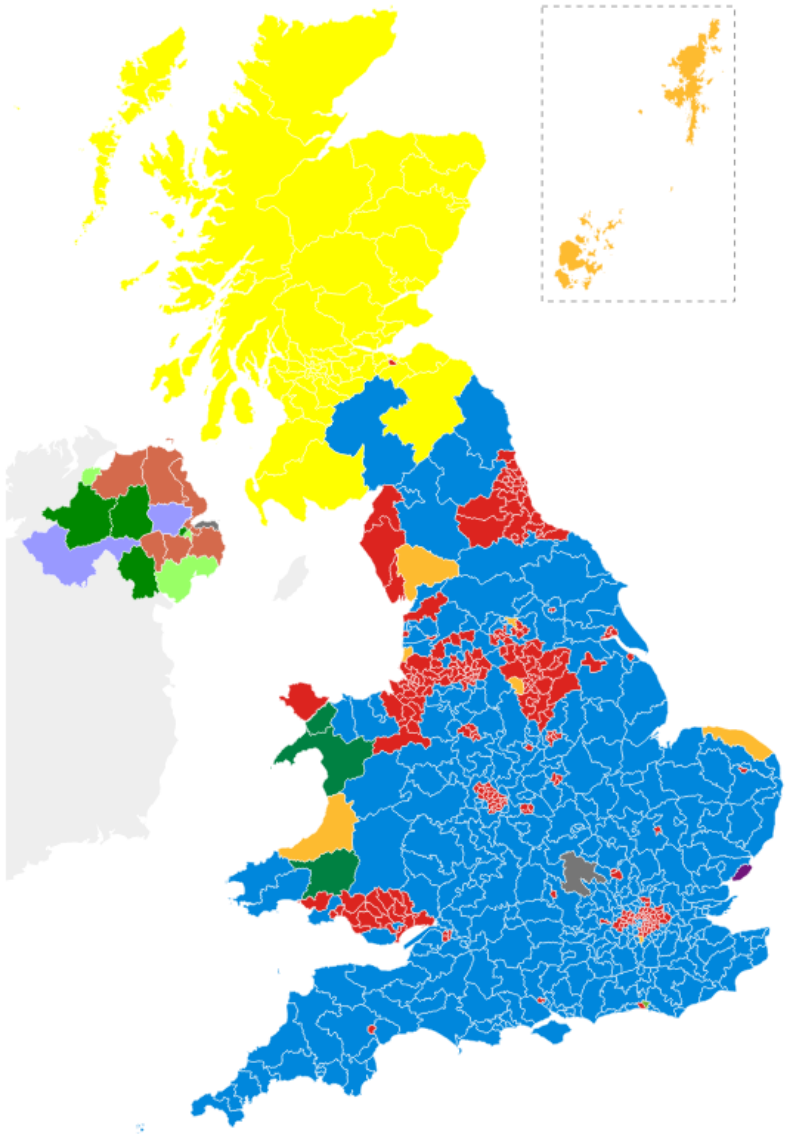
2015 UK General Election Results: Table

Party Standings

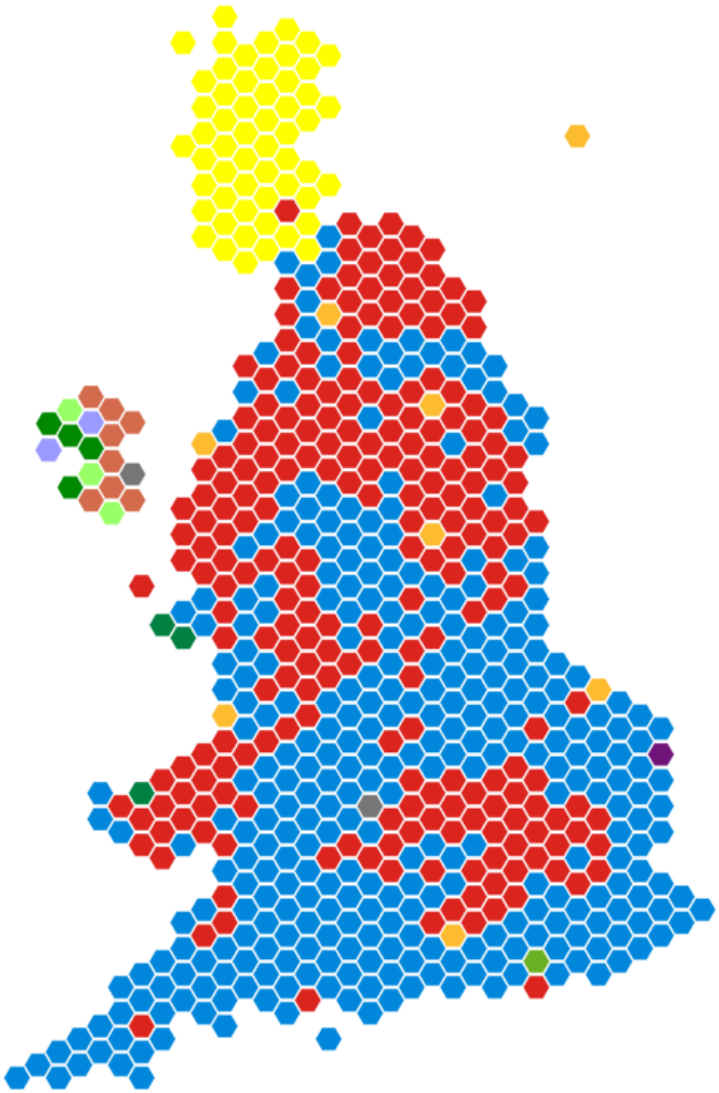
Party	Seats	+/−	Votes	%
Conservative	330	+28 ▲	11,300,109	36.8
Labour	232	-25 ▼	9,347,324	30.5
Scottish National Party	56	+50 ▲	1,454,436	4.7
Liberal Democrat	8	-49 ▼	2,415,862	7.9
Democratic Unionist Party	8	0 —	184,260	0.6
Sinn Fein	4	-1 ▼	176,232	0.6
Plaid Cymru	3	0 —	181,704	0.6
Social Democratic and Labour Party	3	0 —	99,809	0.3
UKIP	1	-1 ▼	3,881,064	12.6
Green	1	0 —	1,156,149	3.8
Alliance	0	-1 ▼	61,556	0.2
Respect the Unity Coalition	0	-1 ▼	9,989	0.0
Other	4	0 —	428,012	1.4

Note: The Speaker, who remains neutral, is included in other

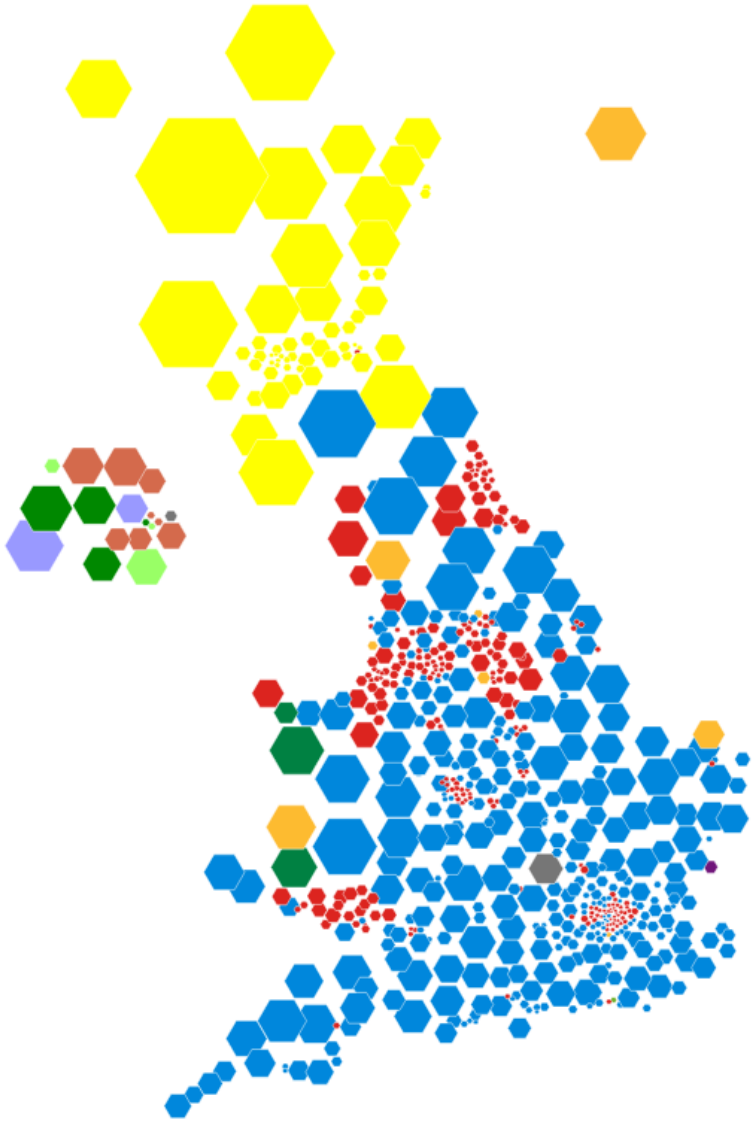
2015 UK General Election Results: Geography Map



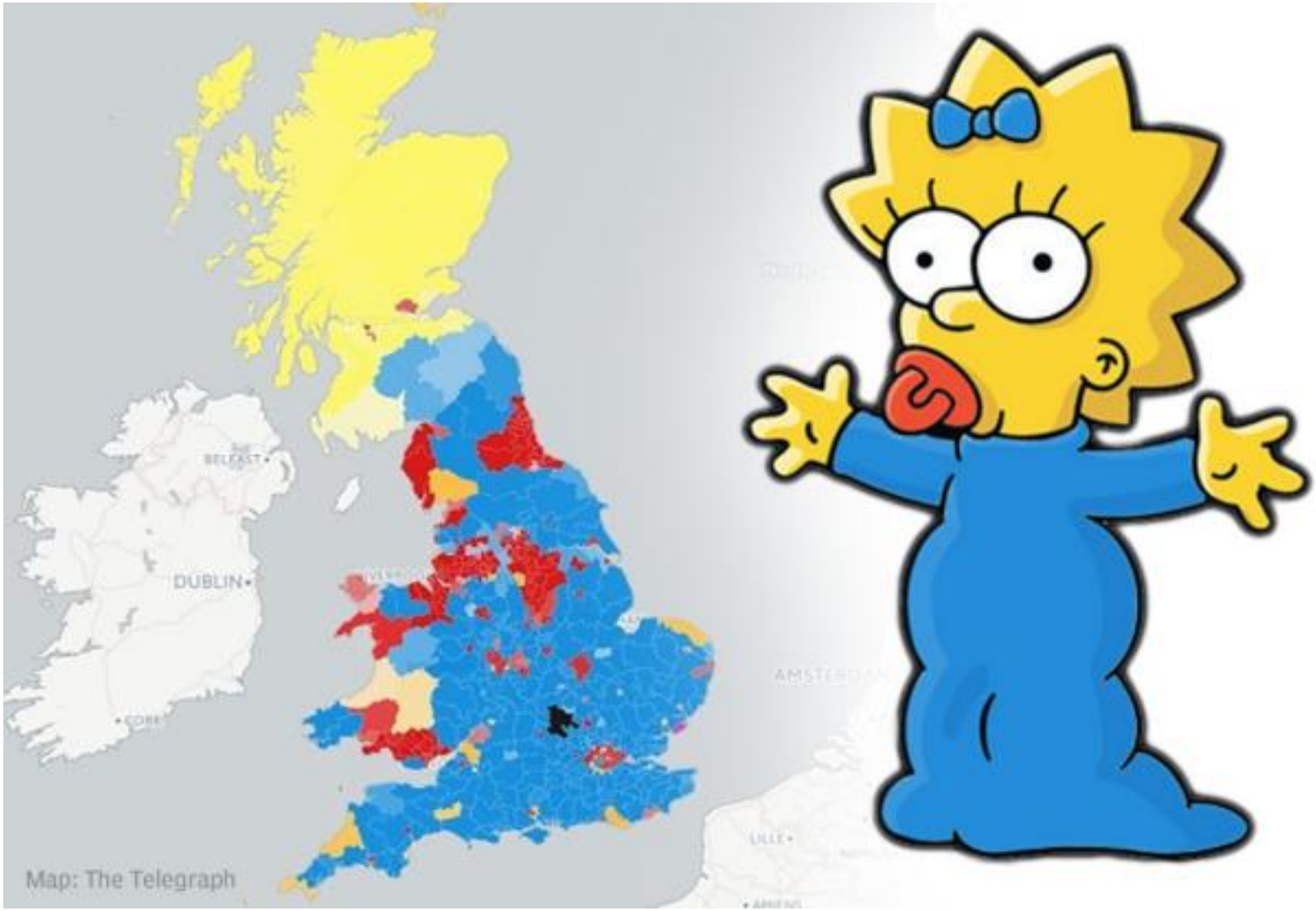
2015 UK General Election Results: Constituency Map



2015 UK General Election Results: Proportional Map



2015 UK General Election Results: Maggie Simpson Map



How Do We Use Data?

1. Description
    - e.g., census demographics, election results, etc.

How Do We Use Data?

1. Description
   - e.g., census demographics, election results, etc.
2. Explanation
   - Testing our beliefs about the way the world works

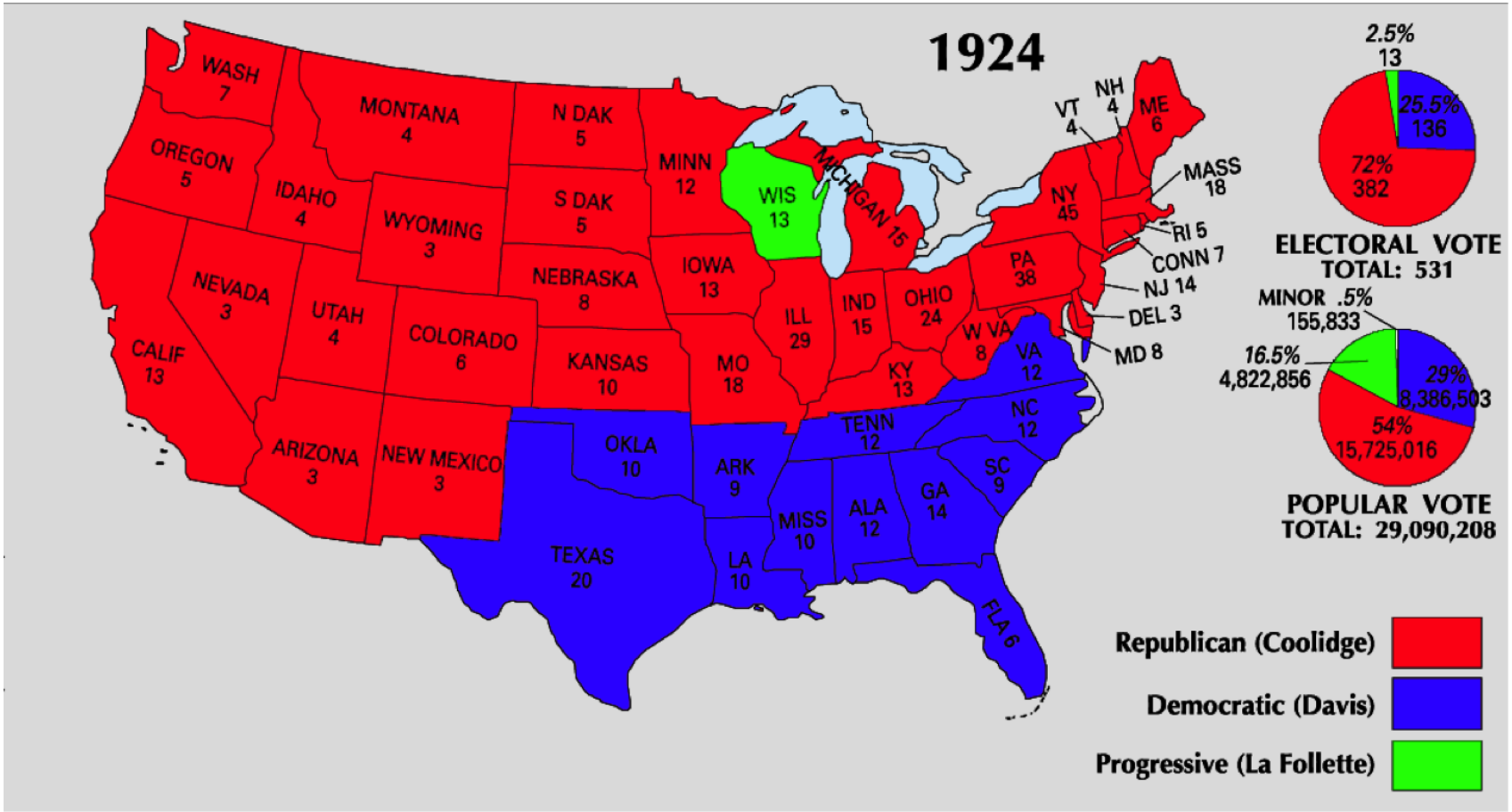
The Theory of Realigning Elections

- Major shift in party support in the electorate
  - Cyclical or generational (i.e., every 30 years)
  - Caused by a major event or issue
  - Witnessed in a 'critical' election

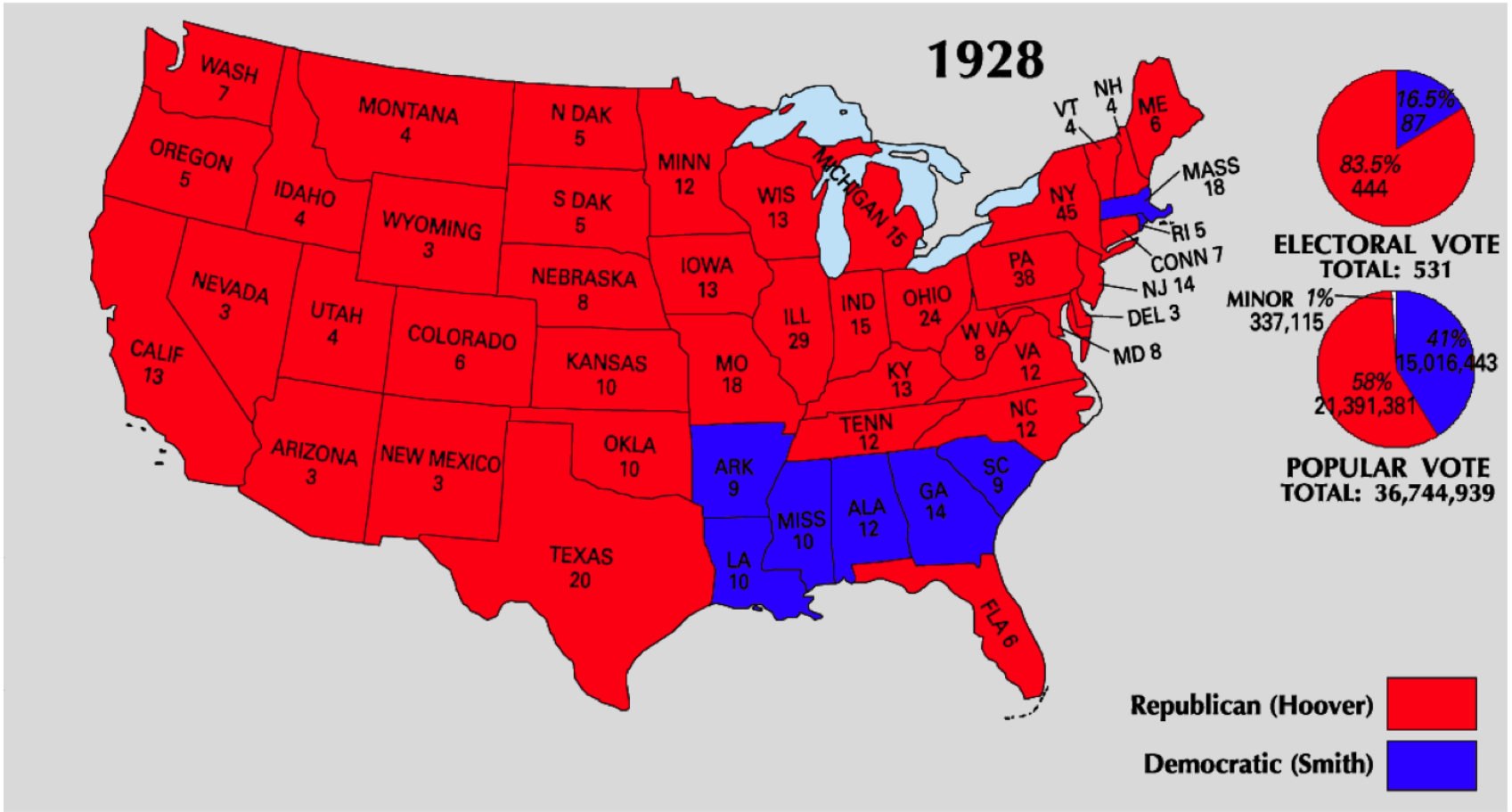
The Theory of Realigning Elections

- Major shift in party support in the electorate
  - Cyclical or generational (i.e., every 30 years)
  - Caused by a major event or issue
  - Witnessed in a 'critical' election
- e.g., 1896 - 1928 U.S. Presidential Elections
  - Republicans won 7 of 9 elections (except '12 & '16 to Wilson)
  - What happened between 1928 and 1932?

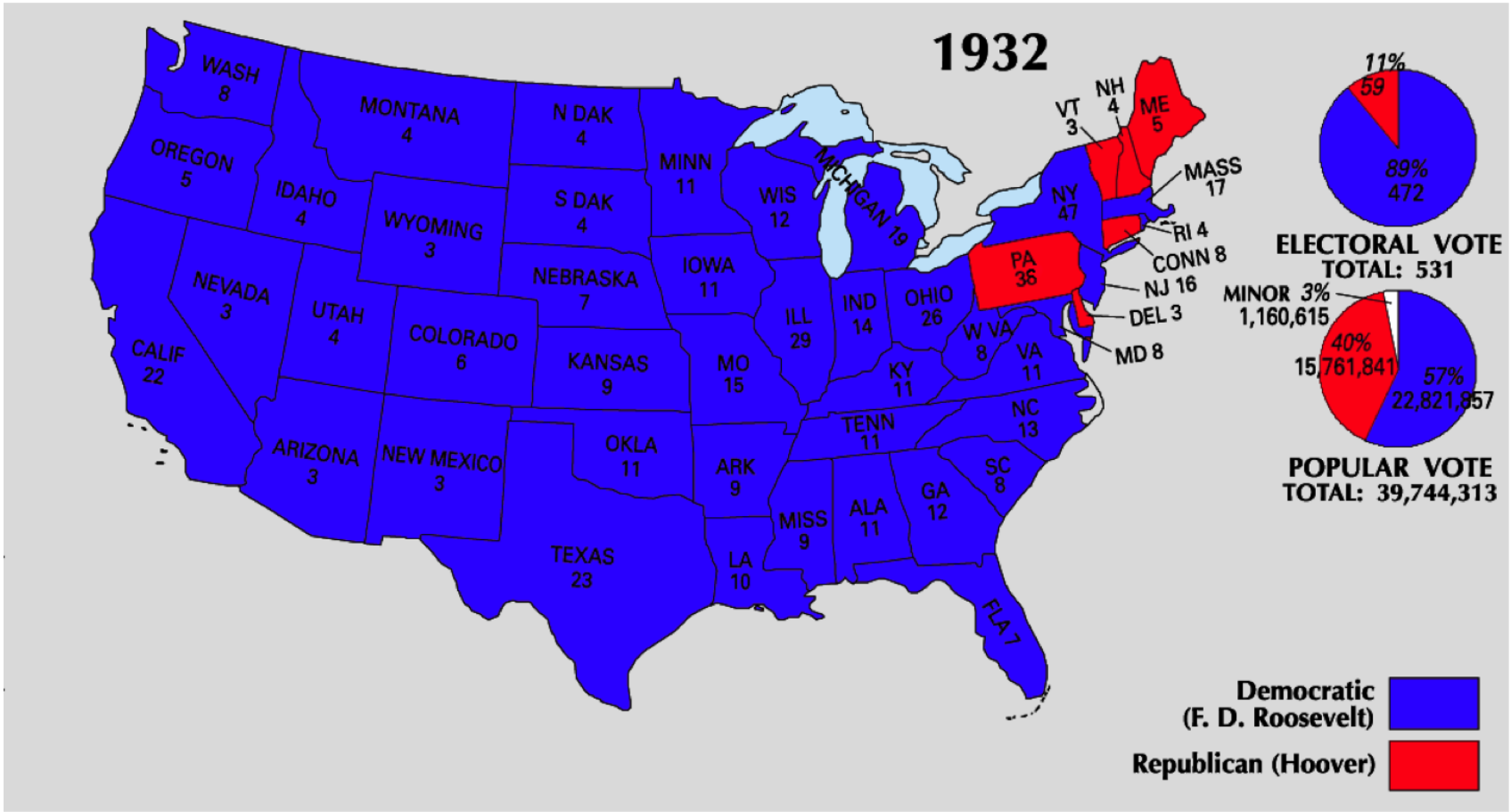
The Theory of Realigning Elections: 1924



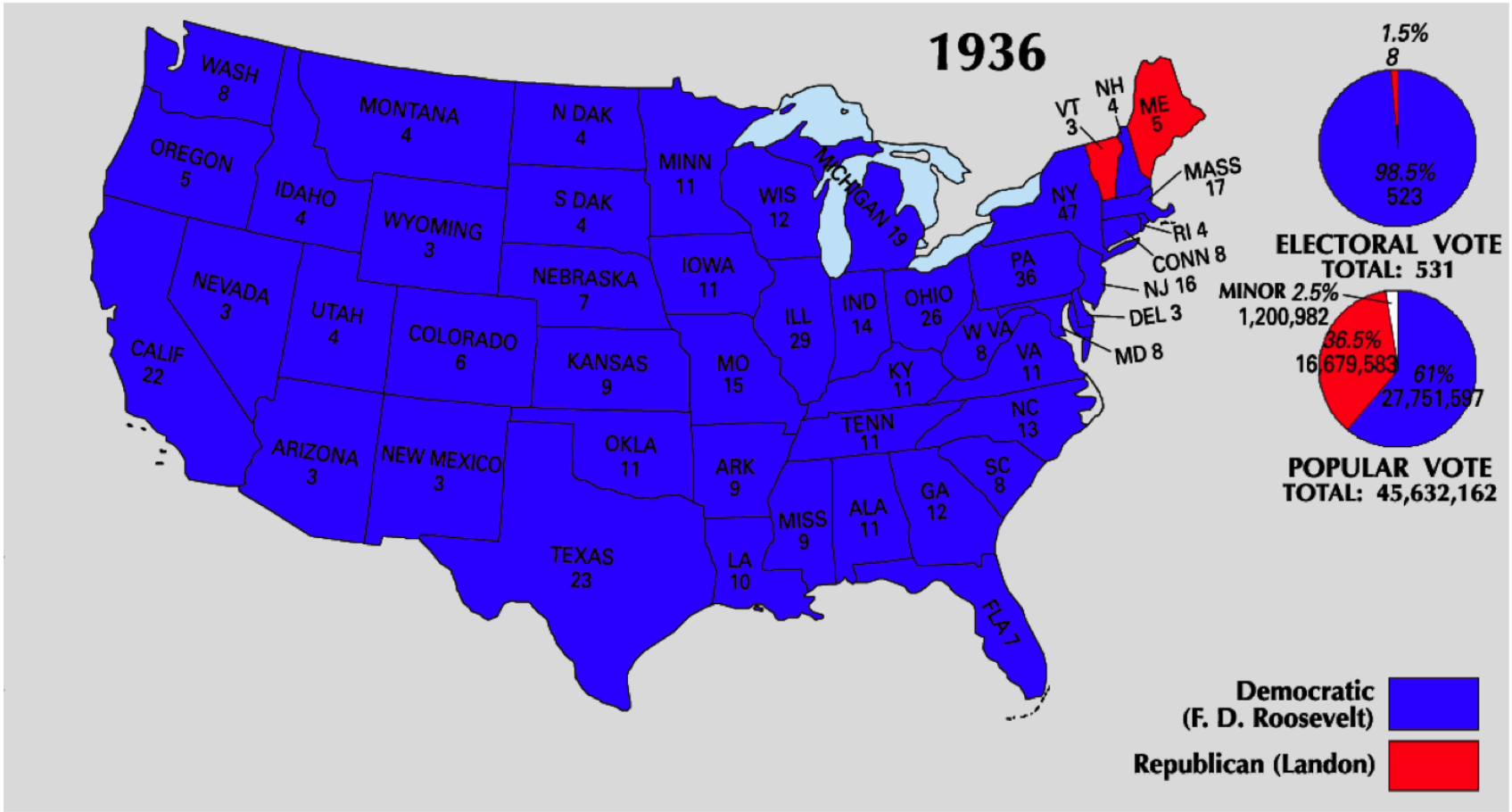
The Theory of Realigning Elections: 1928



The Theory of Realigning Elections: 1932



The Theory of Realigning Elections: 1936



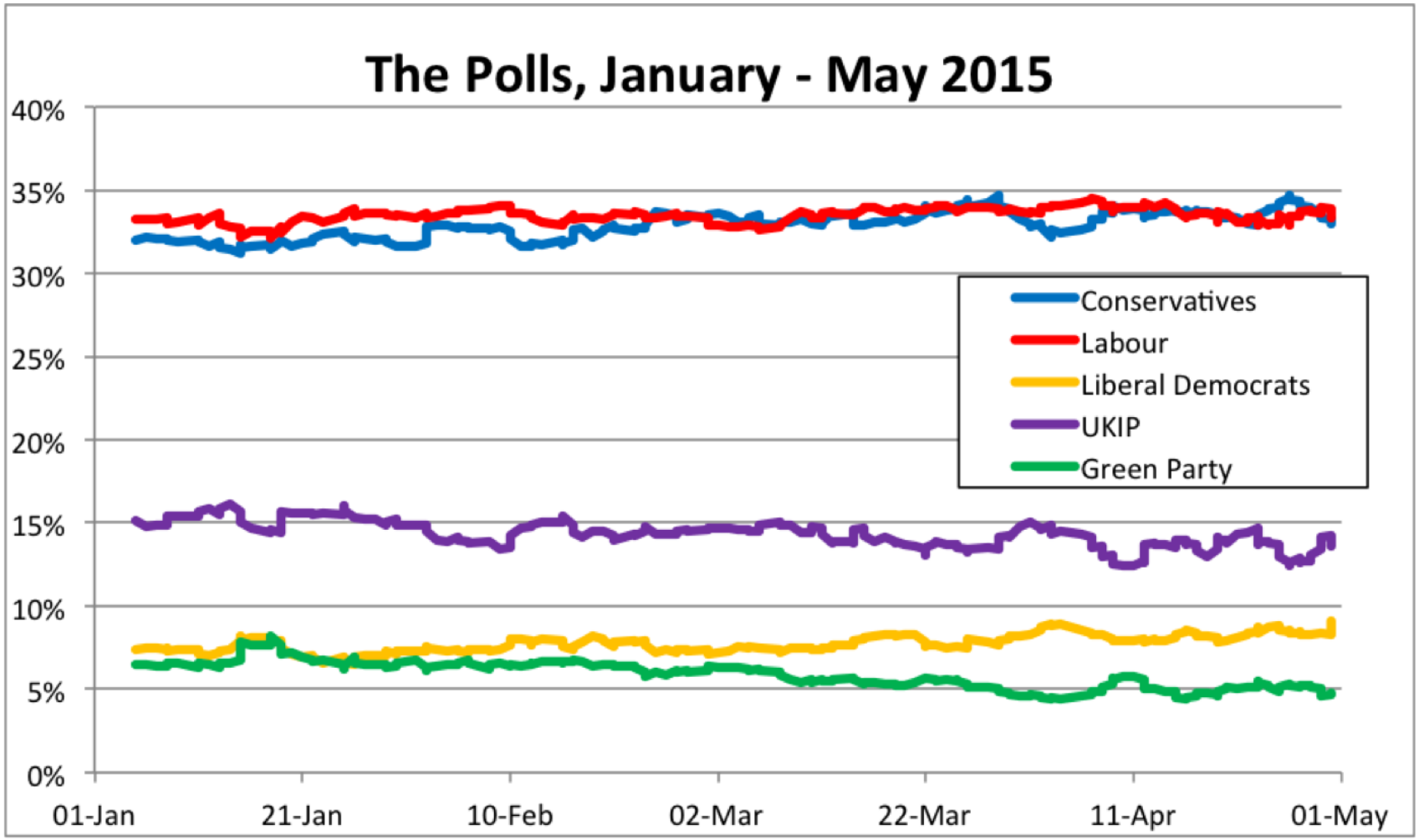
How Do We Use Data?

1. Description
   - e.g., census demographics, election results, etc.
2. Explanation
   - Testing our beliefs about the way the world works

How Do We Use Data?

1. Description
   - e.g., census demographics, election results, etc.
2. Explanation
   - Testing our beliefs about the way the world works
3. Prediction
   - e.g., who will win the upcoming election?

Forecasting the 2015 UK General Election: Polls



Forecasting the 2015 UK General Election: Headlines

Section:GDN BE PaGe:1 Edition Date:150507 Edition:01 Zone:N Sent at 6/5/2015 21:21

Britain's verdict The sharpest comment, analysis and reporting with Owen Jones, Rowena Mason, John Crace Pages 6-11

£1.80 (IR €2.00) Thursday 07.05.15 Published in London and Manchester theguardian.com

Winner of the Pulitzer prize 2014

theguardian

It couldn't be closer

Latest Guardian/ICM poll

35%	35%	11%	9%	3%	7%
Labour	Conservatives	Ukip	Lib Dem	Greens	Others

A late surge in support for Ed Miliband's party will confound David Cameron's aides who predicted a swing to the Tories in the final days Photographs: Oli Scarff/Getty

- Labour gains three points to match Tories
- Greens and Ukip both suffer drop in support
- Parties prepare for coalition talks behind the scenes

Tom Clark and Patrick Wintour

Britain is heading for a second hung parliament in succession with the most drawn-out election campaign since the war apparently ending in near deadlock with Labour and the Conservatives tied at 35% each according to the preliminary results of the final Guardian/ICM poll.

Ed Miliband's party has pulled back three points on ICM's previous campaign poll, published nine days ago, with the Tories remaining unchanged. Previous ICM surveys had reported Conservative leads, ranging from six to two points.

Labour's recovery appears to have been helped by a last-week squeeze at the political fringe: Ukip and the Greens have both slipped back two points, to 11% and 3% respectively. The Liberal Democrats are unchanged on 9%, a buoyant Scottish National party climbs one to a UK-wide score of 5%. Plaid Cymru are on 1%, and other minor parties are also on 1%.

Conservatives were last night still predicting that their dominant message warning of a Miliband government dependent on the votes of the Scottish Nationalists would sway the many undecided voters as they enter the polling booths today.

But on the basis of the ICM survey so far, the long promised swing back to the Conservatives in the final 72 hours before the election – the central pillar of the strategy organised by Lynton Crosby, David Cameron's campaign manager – has not materialised.

The survey is ICM's preliminary prediction poll, with a larger than normal sample size, and will be updated today, after additional interviews conducted late last night are fed into the data.

Conservatives are also hopeful the swing to Labour in the key English marginals will be blunted, pointing to the experience of 1992 and previous opinion polls that have overestimated the Labour share of the vote.

Updating the Guardian's poll projection (an average of all the polls made public) to take account of the new ICM data puts the Conservatives on 274 seats and Labour on 271 – neither anywhere near the 326 required for an absolute majority.

But the huge anti-Tory SNP blue of 52 gives Miliband the stronger position in the battle to negotiate for control of No 10. The cabinet secretary, Sir Jeremy Heywood, is on hand to prepare for talks, but will not press Cameron to leave Downing Street, since it will be his right as prime minister to decide whether to test the opinion of the Commons on a Queens Speech.

Behind the scenes all three parties have been preparing for coalition talks. Labour is hoping on Friday to be able to declare that Cameron cannot command a Commons majority and should resign, challenging Nick Clegg to abandon any thought of a second coalition with him.

Labour sources suggest the closer Cameron is to 280 seats, as opposed to 260, the harder it will be for him to mount a political argument around the illegitimacy of a minority Labour government.

Labour has plans for a big appeal to the Lib Dems to join in building a new coalition to take the country from austerity, but in reality all will turn on the precise distribution of seats between the main parties, something that in turn depends on a few tens of thousands of votes distributed in about 30 marginals.

In a final election rally Ed Miliband insisted the energy in the campaign lay with Labour, saying the Tories had run a negative campaign with nothing to say to working people, nothing to offer for the future – just desperate tactics.

Cameron ended a 36-hour criss-cross tour of key marginals at Hetherington Livestock Mart in Carlisle, where he asked for more time to build a better Britain and said the election would define a generation.

ICM's final survey does not only ask voters about their own voting intention but also about their expectations as to what the rest of the country will do.

Voters' own predictions are less favourable to Labour than the poll headlines suggest. When asked to predict the poll scores, on average, respondents expect Labour to secure only 32, three points less than the 35% they expect for the Tories.

ICM pressed voters on the government they would prefer. The single most popular choice – picked by 25% – was a Conservative overall majority, followed by a Labour majority, which 23% preferred. A Conservative-led administration in which the Tories

Continued on page 2 →

Exclusive Daniel Taylor

FA 'sought to put a limit on England's black players'

The Football Association tried to impose an unofficial quota system on the number of black players an England manager should be allowed to use, according to a new book about racism within the sport.

Pitch Black, written by Amy Onuoha, includes a passage in which one of the sport's anti-racism campaigners claims former England manager Graham Taylor admitted to him he had been summoned by two members of the FA's hierarchy and told "in no uncertain terms" he should not go beyond a certain limit.

There is absolutely no suggestion that Taylor, with a history of working with some of England's leading black players over the past 30 years, adhered to this policy but the book recounts him saying senior figures tried to put pressure on him to keep the team predominantly white.

Taylor is said to have revealed this at a function at Watford FC's ground during the 1999-2000 season when Richie Moran was the guest speaker. Moran, a Birmingham City player in the 1990s who quit the game because of racial abuse he suffered, recalls in the book: "Graham Taylor came up to me and said: 'Look, I'm going to tell you something … I'm never going to admit it, I will be sued for libel.' He said: 'When I was manager of England I was called in by two members of the FA, who I won't name …' I volunteered two names. He said: 'I'm not prepared to say, but I was told in no uncertain terms not to pick too many black players for the national side.'"

This also ties in with a Kick It Out function on its 10th anniversary in 2004, attended by Taylor, when it was reported a former England manager had revealed the same conversation to other guests but declined to go public. The Guardian has spoken to one of the event's organisers, another senior figure in football-related race issues, who said he could confirm Taylor was the manager quoted.

Graham Taylor is said to have admitted he came under pressure to limit the number of black players he selected for England squad

Taylor has been made aware of the book and told the Guardian he could not specifically remember the conversations with Moran or one at the Kick It Out function. "That is not me trying to evade it – and it also doesn't mean I didn't say it – but if anyone looks at my record with club and country it would be obvious I didn't follow what was apparently said. If anyone looks at my record, I could never be accused of blocking the way for any black player."

Taylor later went on BBC Radio 5 Live to say he had never received orders not to pick black players and that a private conversation – if there had been one – should not have been made public. The 70-year-old said he would consider taking legal advice about the publisher and author.

Onuoha's book, officially launched on Friday, makes the same point but the claims potentially raise serious questions about the attitudes of some FA figures during a period of the game – Taylor managed England from 1990 to 1993 – when the sport was supposed to be leaving behind more obvious elements of racism.

The FA has been notified about the claim and chosen not to comment. No members of the current FA board were in charge at the time of the allegation.

Moran told the Guardian yesterday: "I have a very vivid memory of the conversation. My then girlfriend said to me after he'd told me: 'Who was that?'

"I said: 'It's the former England manager.' She said: 'Well why did he tell you that then?' I'm not saying for one moment that Graham Taylor had any intentions … all I'm saying is that is a conversation I had with him."

THE SCOTSMAN

THURSDAY 7 MAY 2015 SCOTLAND'S NATIONAL NEWSPAPER £1.40 SUBSCRIPTION PRICE £1.12 WWW.SCOTSMAN.COM

24 pages of Property inside today PULL-OUT

Deila says no big-name summer sell-off SPORT, BACK PAGE

Staff in Falkirk prepare ballot boxes for this morning, as voters decide who will govern the country in the most closely fought election in a generation

'Doctors can win battle with superbugs'

DOCTORS can reverse superbugs' resistance to antibiotics using mathematical formulas, reducing the chance of a major epidemic of infections which could kill millions.

Scientists from the University of California and the American University in Washington DC have found a way of combining antibiotics using "Time Machine" software. The technique allows them to rewind the evolution of bacteria and reverse resistance to treatments.

Dr Miriam Barlow, of the University of California, said: "This work shows that there is still hope for antibiotics if we use them intelligently."

FULL STORY, PAGE 9

Sainsbury's profits down after 10 years

SAINSBURY'S has reported its first profits drop in a decade as "unprecedented" changes in the retail market continue.

Britain's third-largest supermarket posted underlying profits down 147 per cent at £681 million, while pre-tax losses – including a writedown on land on which it will not now build – hit £72m as it warned of continuing market pressures in years ahead.

BUSINESS, PAGE 30

44 SCOTLAND STREET AT THE CANNY MAN'S Page 22

Final election polls put it neck-and-neck

DAVID MADDOX
POLITICAL EDITOR

Latest figures have Labour and Tories equal in UK race and big gains for SNP

FEVER-OF-ELECTION polls put absolutely nothing between Labour and the Conservatives last night as the country prepared to vote in what promises to be the most tightly fought political battle in a generation.

An ICM UK poll put both Labour and the Tories on 35 per cent with Ed Miliband's party pulling back three points from the last ICM poll and the Conservatives' score unchanged.

A final YouGov poll suggested they will both win 276 seats.

However, in Scotland, a Panelbase poll put the Nationalists on 48 per cent of the vote, ahead of Labour on 26 per cent – support which, if transferred into seats, would give the SNP a landslide, gaining all but six of the country's 59 seats.

If that turned out to be the case, it would put First Minister Nicola Sturgeon in a potentially powerful position as the SNP would likely be the third-biggest party in Westminster.

Yesterday Ms Sturgeon made a last-ditch appeal to Labour leader Mr Miliband to talk to her if his party and the SNP can hold a majority between them.

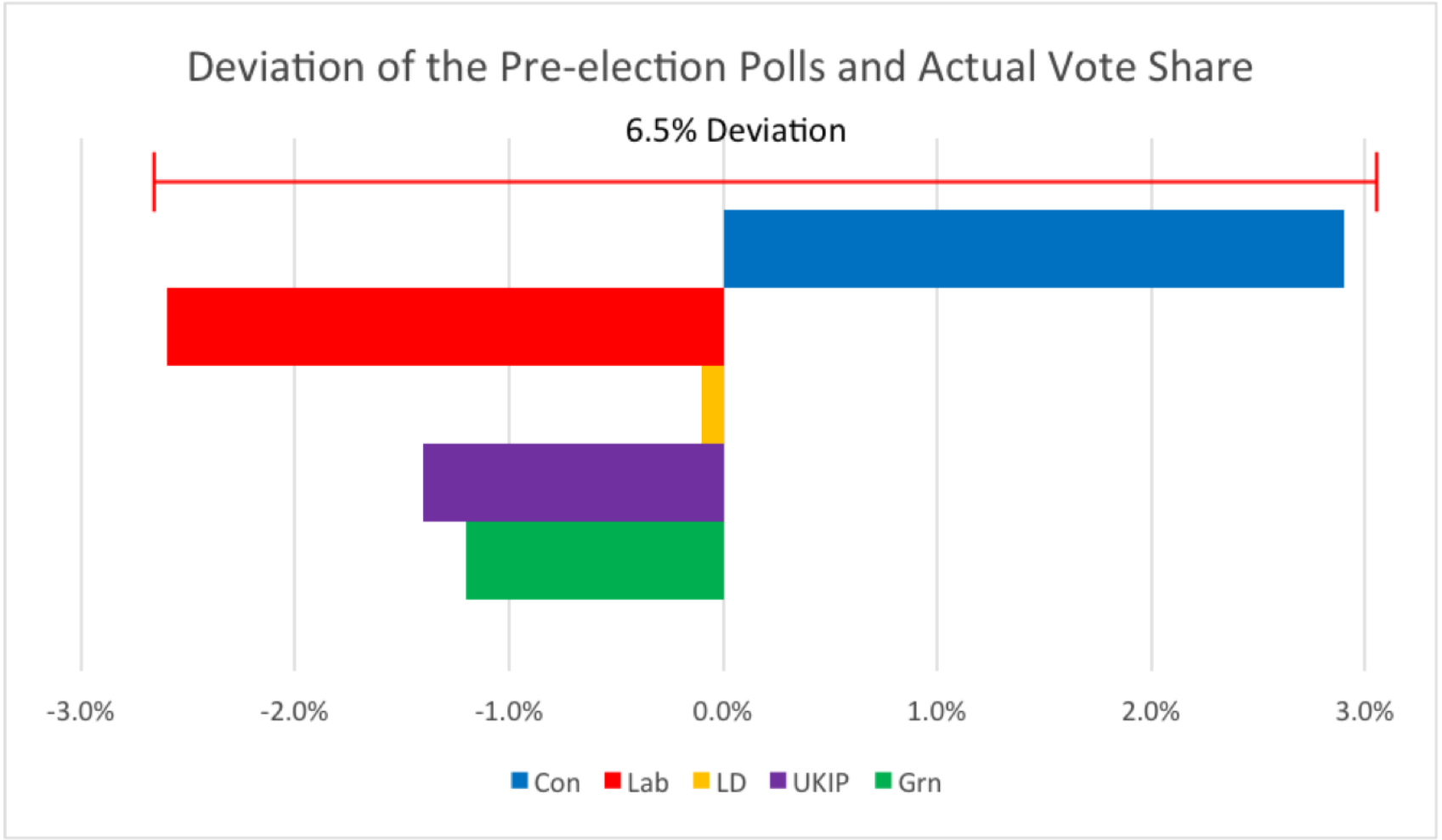
While Ms Sturgeon has said she wants

CONTINUED ON PAGE 4

Forecasting the 2015 UK General Election: Models

	Five Thirty Eight	May 2015	Elections Etc	The Guardian	ACTUAL
Conservatives	278	273	285	273	**331**
Labour	267	268	262	273	**232**
Liberal Democrats	27	28	25	27	**8**
SNP	53	56	53	52	**56**
UKIP	1	2	3	3	**1**
Other	24	23	23	22	**22**
RESULT	**HUNG**	**HUNG**	**HUNG**	**HUNG**	**CONSERVATIVE**

Forecasting the 2015 UK General Election: Oops!



Measuring the World: Variables

- A *variable* is a quantitative measurement

Measuring the World: Variables

- A *variable* is a quantitative measurement
- Two types of variables
  1. Dependent (a.k.a. outcome) variable
     - In politics: e.g., vote choice, political attitudes, etc.
     - In criminology: e.g., crime rates, incarceration, etc.
     - In management: e.g., organizational performance, consumer behaviour, etc.

Measuring the World: Variables

- A *variable* is a quantitative measurement
- Two types of variables
  1. Dependent (a.k.a. outcome) variable
     - In politics: e.g., vote choice, political attitudes, etc.
     - In criminology: e.g., crime rates, incarceration, etc.
     - In management: e.g., organizational performance, consumer behaviour, etc.
  2. Independent (a.k.a. predictor) variable(s)
     - e.g., gender, age, race/ethnicity, party affiliation, etc.

Variables

- A *dependent variable* is what we want to explain
  i.e., its variation *depends* on one or more variables

Variables

- A *dependent variable* is what we want to explain
  i.e., its variation *depends* on one or more variables
  - In equations, generally referred to as y_i

Variables

- A *dependent variable* is what we want to explain
  i.e., its variation *depends* on one or more variables
  - In equations, generally referred to as y_i
- *Independent variables* are used to explain a dependent variable
  i.e., it will be used to predict the dependent variable
  - In equations, generally referred to as x_i

  Thus, $x_i \rightarrow y_i$ (x_i predicts or causes y_i)

Levels of Measurement

1. Nominal Variable (a.k.a. Categorical Variable)
   - Attributes are defined by NAME only (i.e., no inherent ordering to the categories)
   - e.g., country, gender, religion, etc.

Levels of Measurement

1. Nominal Variable (a.k.a. Categorical Variable)
   - Attributes are defined by NAME only (i.e., no inherent ordering to the categories)
   - e.g., country, gender, religion, etc.
2. Ordinal Variable
   - Categories are properly ORDERED
   - But distance between values is not consistent
   - e.g., Likert scales, polling questions, education, etc.

Levels of Measurement

1. Nominal Variable (a.k.a. Categorical Variable)
   - Attributes are defined by NAME only (i.e., no inherent ordering to the categories)
   - e.g., country, gender, religion, etc.
2. Ordinal Variable
   - Categories are properly ORDERED
   - But distance between values is not consistent
   - e.g., Likert scales, polling questions, education, etc.
3. Continuous or Pseudo-Continuous Variable
   - Values are ordered and distance is consistent
   - e.g., time, fiscal spending, etc.
   - Interval scale (e.g., IQ test has no real 0)
   - Ratio scale (e.g., Kelvin, where zero is meaningful)

Descriptive Statistics: The Mean

- Sum of values divided by total # of values

$$\bar{y} = \frac{y_1 + y_2 + y_3 + \ldots y_n}{n}$$

Mean

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

The Mean: An Example

- 2012 London Olympics: Men's 100m Final

Rank	Athlete	Country	Result
1	Usain BOLT	JAM	9.63
2	Yohan BLAKE	JAM	9.75
3	Justin GATLIN	USA	9.79
DQ	Tyson GAY	USA	9.80
5	Ryan BAILEY	USA	9.88
6	Churandy MARTINA	NED	9.94
7	Richard THOMPSON	TTO	9.98
8	Asafa POWELL	JAM	11.99

The Mean: An Example

- What is the mean 100m time?
- Sum of values divided by total # of values

$$\bar{y} = \frac{9.63 + 9.75 + 9.79 + 9.80 + 9.88 + 9.94 + 9.98 + 11.99}{8}$$

$$= \frac{80.76}{8}$$

$$= 10.095$$

An Outlier: Asafa Powell

Descriptive Statistics: The Median

- The middle value in a distribution
- Less sensitive to outliers (i.e., extreme values)
- To determine:
  - Order the values from low to high
  - For odd # of values, median is middlemost value
  - For even # of values, median is mean of two middle values

The Median: An Example

- What is the median 100m time?
- Recall, order the values and select middlemost value

  Order the race times: 9.63, 9.75, 9.79, 9.80, 9.88, 9.94, 9.98, 11.99

  $$\text{Median} = \bar{y} = \frac{9.80 + 9.88}{2} = 9.84$$

Comparing the Mean & Median: YouGov Poll

YouGov

YouGov

Sample	1000 Adult Interviews
Conducted	March 27 - 29, 2015
Margin of Error	±3.9%

1. Do you believe in love at first sight?

Yes	51%
No	35%
Not sure	14%

Comparing the Mean & Median: YouGov Poll

2. Do you believe there is such a thing as a soulmate?

Yes	69%
No	18%
Not sure	13%

Comparing the Mean & Median: YouGov Poll

3. Do you think a person has only one soulmate or do you think a person can have more than one?

Asked of people who believe there is such a thing as a soulmate

Only one soulmate	25%
More than one soulmate	33%
Not sure	10%
Does not believe in or isn't sure about soulmates	31%

Comparing the Mean & Median: YouGov Poll

4. How many soulmates do you think a person can have?
Asked of people who believe there is such a thing as a soulmate

mean	3,741,057
median	2

Descriptive Statistics: The Variance

- Sum of squared deviations from the mean divided by total # of values

$$s^2 = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \ldots (y_n - \bar{y})^2}{n}$$

Variance

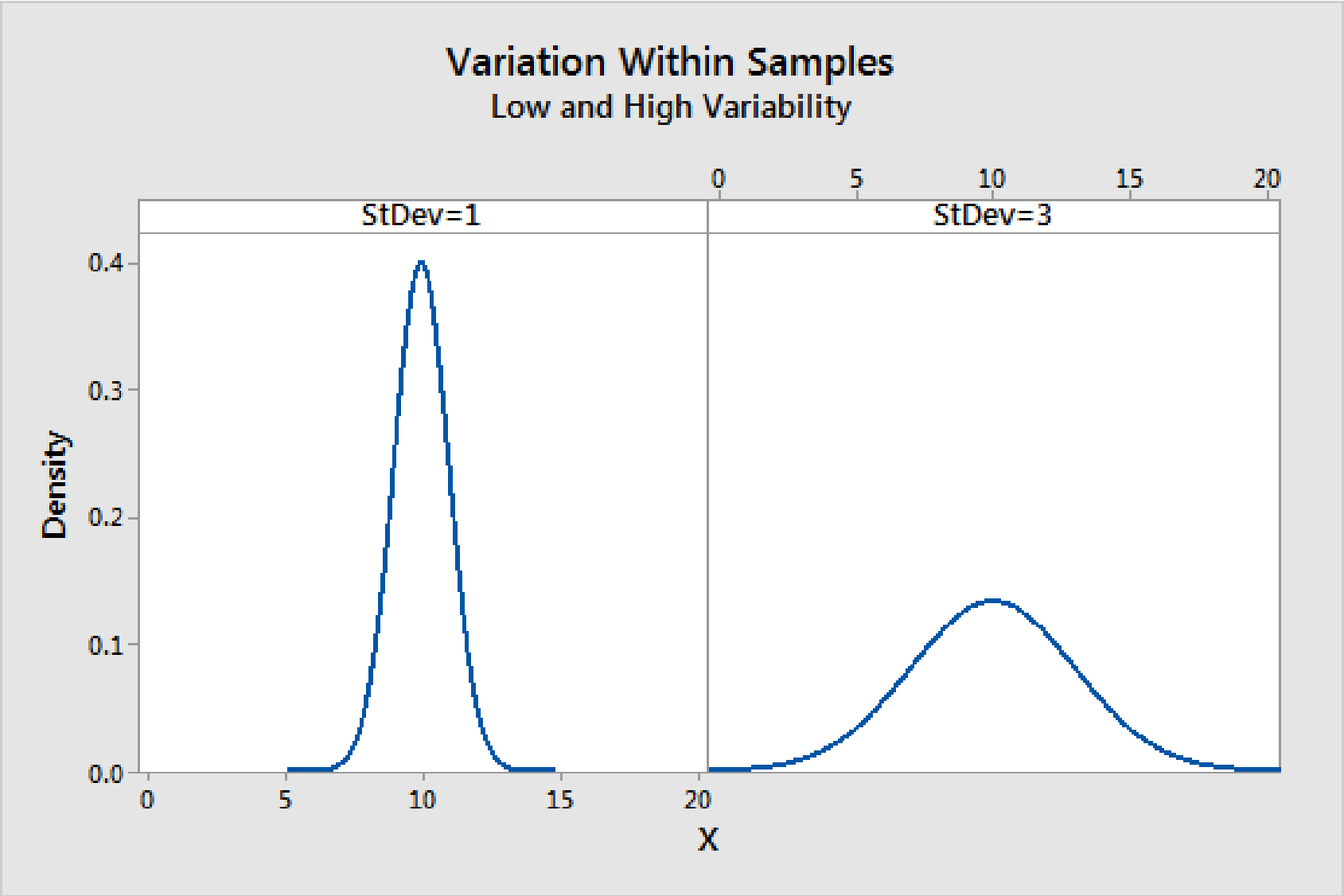
$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Descriptive Statistics: The Standard Deviation

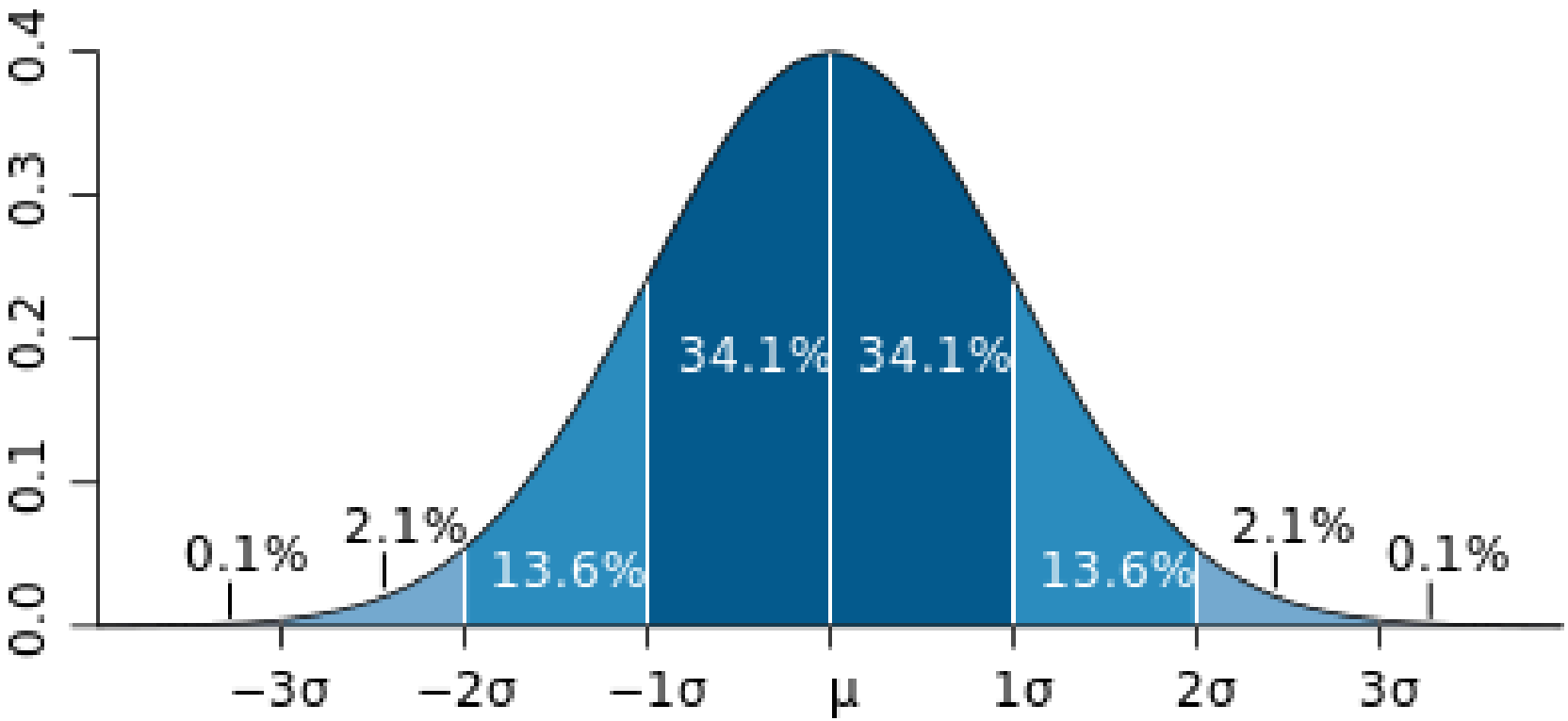
Standard Deviation

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Low and High Variation



Standard Deviation Distribution



It's All Greek to Me

GREEK ALPHABET

By Ben Crowder • bencrowder.net • Last modified 2 May 2012

Αα	Ββ	Γγ	Δδ	Εε	Ζζ
ALPHA [a]	**BETA** [b]	**GAMMA** [g]	**DELTA** [d]	**EPSILON** [e]	**ZETA** [dz]
ἄλφα	*βῆτα*	*γάμμα*	*δέλτα*	*ἒ ψιλόν*	*ζῆτα*
Ηη	Θθ	Ιι	Κκ	Λλ	Μμ
ETA [ɛː]	**THETA** [tʰ]	**IOTA** [i]	**KAPPA** [k]	**LAMBDA** [l]	**MU** [m]
ἦτα	*θῆτα*	*ἰῶτα*	*κάππα*	*λάμβδα*	*μῦ*
Νν	Ξξ	Οο	Ππ	Ρρ	Σσς
NU [n]	**XI** [ks]	**OMICRON** [o]	**PI** [p]	**RHO** [r]	**SIGMA** [s]
νῦ	*ξεῖ*	*ὂ μικρόν*	*πεῖ*	*ῥῶ*	*σῖγμα*
Ττ	Υυ	Φφ	Χχ	Ψψ	Ωω
TAU [t]	**UPSILON** [ʉ]	**PHI** [pʰ]	**CHI** [kʰ]	**PSI** [ps]	**OMEGA** [ɔː]
ταῦ	*ὗ ψιλόν*	*φεῖ*	*χεῖ*	*ψεῖ*	*ὦ μέγα*