Bush 631-607: Quantitative Methods

Lecture 5 (09.28.2021): Measurement vol. II

Rotem Dvir

The Bush school of Government and Public Policy

Texas A&M University

Fall 2021

What is today's plan?

- More on measurement.
- Latent concepts.
- Correlation.
- Visuals: scatterplots.
- Clustering.
- R work: scatterplot, subset(), grouping, kmeans()

Measurement

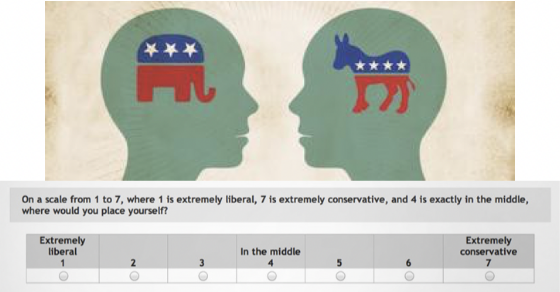
Why?

- Social science: develop and test causal theories.
- Leader background and conflict behavior.
- Minimum wage and levels of full-time employment?
- Concepts: level of unemployment, leader background, public approval.

How?

Measures - the context of theoretical concepts

Measureing ideology



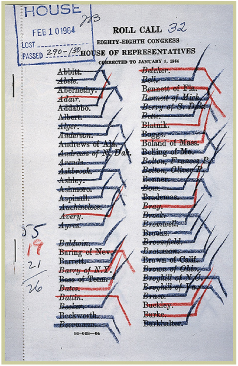
Measurement models:

- Summarize data.
- Learn about human behavior.

Measuring ideology

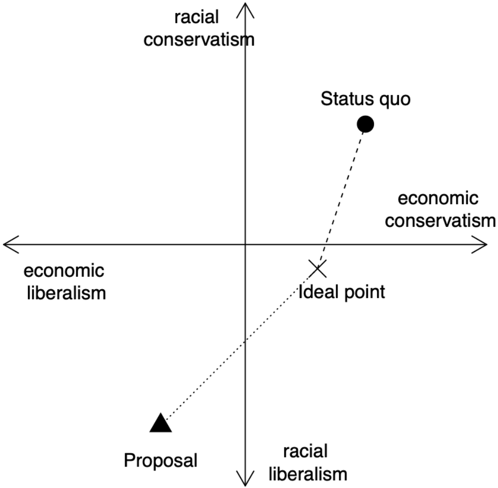
Legislators measurement model: congress roll-call votes

Infer from behavior: voting → orientation.



Ideology in US Congress

Spatial voting: voting and political ideology



Complex measurement

Latent concepts:

- Hard to measure.
- Variation in definitions.
- Democracy: the polity debate.
- Ideology: representative votes?

Other suspects:

- Terrorism: which violent events are terrorism?
- Resolve: how resolve is the president?

What is terrorism?

Researchers $\rightarrow$ objective measures:

- Identity: perpetrators and victims.
- Population-wide psychological effects.
- Clear political objective.

The Public?

You tell me

Public views of terrorism?

Huff and Kertzer (2018):

- Objective: 'facts on the ground'.
- Subjective: 'who and why?'

The Method: Conjoint experiment

- No control group.
- Multiple treatments.
- Outcome: is it terrorism? (yes/no)
- How each factor contributes to viewing an incident as terrorism?

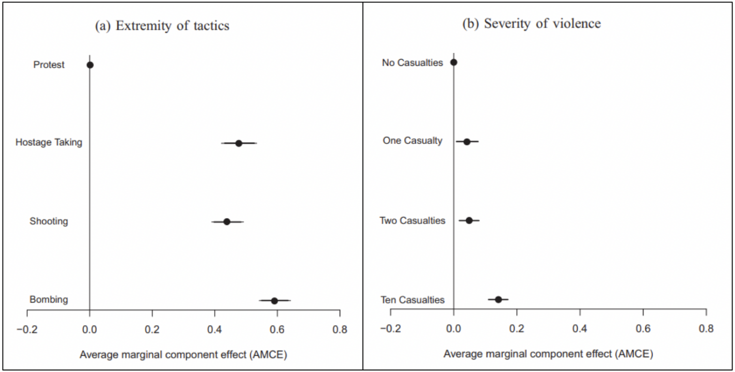
Conjoint experiment: Terrorism

Scenario 1
The incident: shooting
The incident occurred in a church in a foreign democracy with a history of human rights violation
Two individuals died.
The shooting was carried by a Muslim individual with history of mental illness.
News suggest the individual had ongoing personal dispute with one of the targets

Scenario 2
The incident: bombing
The incident occurred in a police station in a foreign dictatorship.
No fatalities reported.
The bombing was carried by a Muslim organization.

News suggest the group was motivated by the goal of overthrowing the government.

Objective path: results



Subjective path: results

Figure 5 Social Categorization Effects

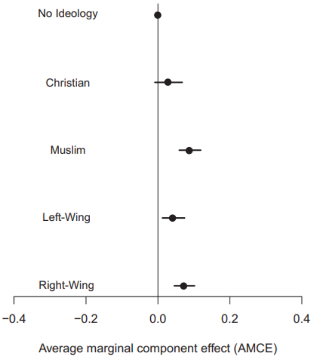
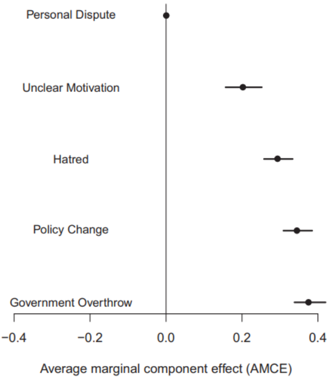


Figure 6 Motive Attribution Effects



Terrorism data

Type: event data

A lot of resources:

- GTD - START (Maryland).
- Individuals radicalization (PRIUS) - START (Maryland).
- Episodes of political violence (1946-2017) (Vienna, Austria).
- Suicide terrorism - CPOST (Chicago)
- List (Link)

Terrorism data

Global Terrorism Database (GTD):

- Time frame: 1970-2019.
- Events: International & domestic terrorism.
- Scope: over 100,000 cases.
- Sources: open source media.

Problem(s)?

- Events data $\rightarrow$ news sources.
- Temporal: less work prior to 1970.
- Biased and Selective reporting: strategic, sensational events.
- Errors in measurement.
- Measures matter - democracy and frequency of incidents (polity, strategic reporting).

Latent concept: Resolve

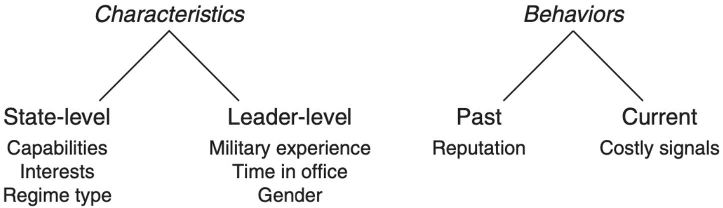
	Country A	Country B
Government	The country is a democracy	The country is a democracy
Interests in the dispute	Experts describe the country's stakes in the dispute as high.	Experts describe the country's stakes in the dispute as high.
Leader background	The leader recently took office; he has served in the military briefly.	The leader recently took office; she had a long career in the military.
Foreign relations	The country is an ally of the United States.	The country is an adversary of the United States.
Previous behavior in international disputes	The last time this country was involved in an international dispute, it initiated the crisis by issuing a public threat to use force against an adversary of the United States, but ultimately backed down. At the time, the country was led by a different leader than the one in the current dispute.	The last time this country was involved in an international dispute, it initiated the crisis by issuing a public threat to use force against an adversary of the United States, and stood firm throughout the crisis. At the time, the country was led by a different leader than the one in the current dispute.
Current behavior	In the current crisis, the country has yet to make any statements or carry out any actions.	In the current crisis, the country has made a public threat that they will use force if the other country does not back down.
Military Capabilities	The country does not have a very powerful military	The country has a very powerful military

In disputes like theses, countries either back down or stand firm.
If you had to choose between them, which of the two countries is more likely to *stand firm* in the current dispute?

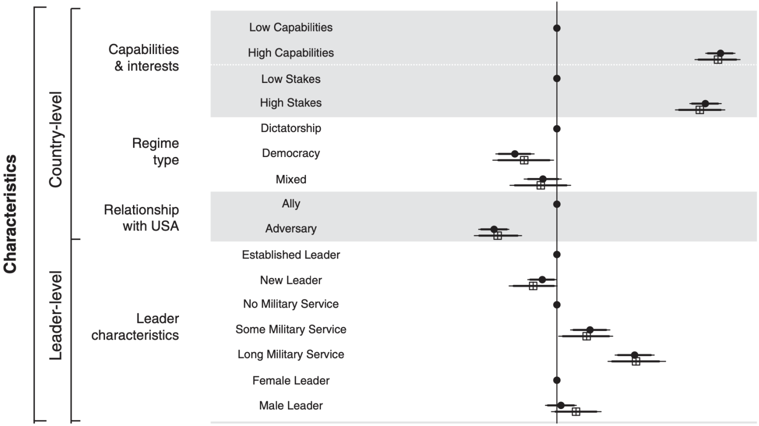
Country A	Country B
○	○

What is resolve?

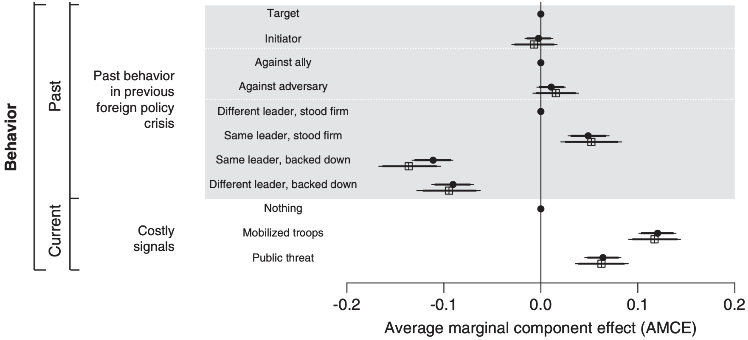
Two paths:



Results


Characteristics
Country-level
Leader-level
Capabilities & interests
Regime type
Relationship with USA
Leader characteristics
Low Capabilities
High Capabilities
Low Stakes
High Stakes
Dictatorship
Democracy
Mixed
Ally
Adversary
Established Leader
New Leader
No Military Service
Some Military Service
Long Military Service
Female Leader
Male Leader

Results


Behavior
Past
Current
Past behavior in previous foreign policy crisis
Costly signals
Target
Initiator
Against ally
Against adversary
Different leader, stood firm
Same leader, stood firm
Same leader, backed down
Different leader, backed down
Nothing
Mobilized troops
Public threat
-0.2
-0.1
0.0
0.1
0.2
Average marginal component effect (AMCE)

Complex concepts & measurement

What's the bottom-line?

- Latent concepts: democracy, ideology, terrorism, resolve.
- Tricky measurement.
- More ways to measure: resolve $\rightarrow$ rival reciprocate in crisis.

How to improve measures?

- Theoretical grounding.
- Replications.

Bivariate Relationships

Summarize relationship b-w 2 variables

Liberal-conservative ideology: Economy & Race

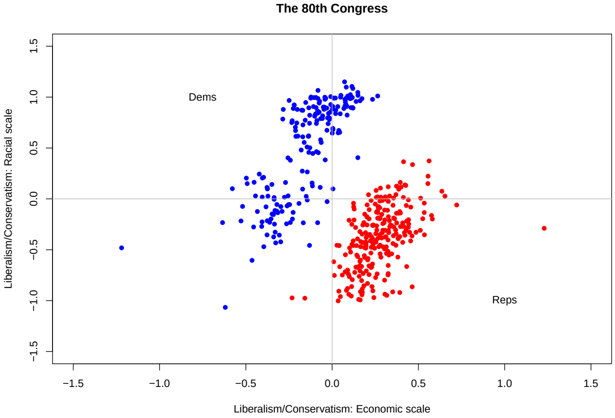
```
head(congress)
```

```
##   congress district   state    party       name dwnom1 dwnom2
## 1       80        0     USA Democrat      TRUMAN -0.276  0.016
## 2       80        1 ALABAMA Democrat  BOYKIN   F. -0.026  0.796
## 3       80        2 ALABAMA Democrat   GRANT   G. -0.042  0.999
## 4       80        3 ALABAMA Democrat ANDREWS   G. -0.008  1.005
## 5       80        4 ALABAMA Democrat   HOBBS   S. -0.082  1.066
## 6       80        5 ALABAMA Democrat   RAINS   A. -0.170  0.870
```

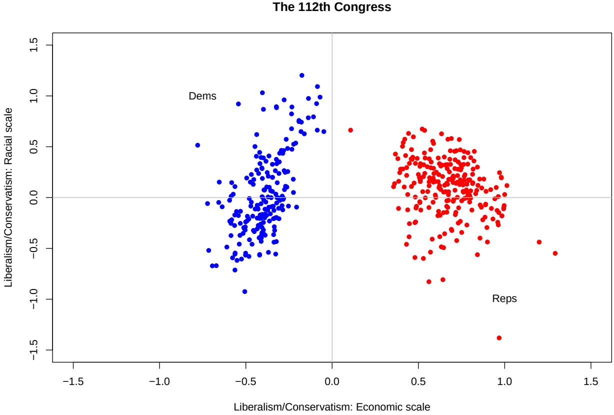
Back to visuals

Scatter plot

- Visualize relationship between 2 variables.
- Numeric/continuous values.

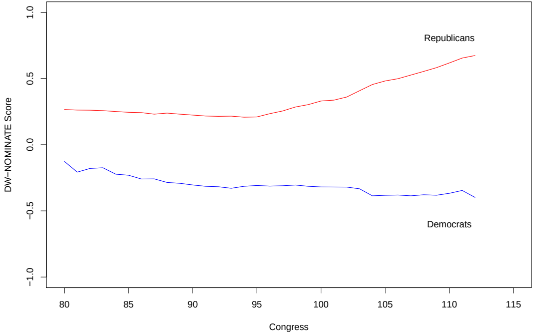


Congress ideology in the 21st century



Congress ideology: time trend

```r
dem.med <- tapply(dem$dwnom1, dem$congress, median)
rep.med <- tapply(rep$dwnom1, rep$congress, median)

plot(names(dem.med), dem.med, col = "blue", type = "l",
     xlim = c(80,115), ylim = c(-1,1), xlab = "Congress",
     ylab = "DW-NOMINATE Score")
lines(names(rep.med), rep.med, col = "red")
text(110, -0.6, "Democrats")
text(110,0.8, "Republicans")
```



‘International’ Ideology

UN $\rightarrow$ International institution.

Voting patterns $\rightarrow$ countries orientation/ideology.

UN voting data (1946-2012)

```
dim(mydata)
```

```
## [1] 9120    6
```

```
summary(mydata)
```

```
##       Year        CountryAbb         CountryName          idealpoint
##  Min.   :1946   Length:9120        Length:9120        Min.   :-2.6552
##  1st Qu.:1972   Class :character   Class :character   1st Qu.:-0.6406
##  Median :1987   Mode  :character   Mode  :character   Median :-0.1644
##  Mean   :1985                                         Mean   : 0.0000
##  3rd Qu.:2001                                         3rd Qu.: 0.7968
##  Max.   :2012                                         Max.   : 3.0144
##
##    PctAgreeUS      PctAgreeRUSSIA
##  Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.1395   1st Qu.:0.5053
##  Median :0.2400   Median :0.6567
##  Mean   :0.2960   Mean   :0.6219
##  3rd Qu.:0.3902   3rd Qu.:0.7424
##  Max.   :1.0000   Max.   :1.0000
##  NA's   :1        NA's   :5
```

Global ideologies

Voting with US → measure of foreign policy similarity.

Similar FP → similar global orientation.

```
# Tidyverse approach to data management
# Arrange by year, calculate mean for US / Russia voting
annual.agree <- mydata %>%
  group_by(Year) %>%
  summarize(us.agree = mean(PctAgreeUS, na.rm = T),
            ru.agree = mean(PctAgreeRUSSIA, na.rm = T))

head(annual.agree)
```

```
## # A tibble: 6 x 3
##    Year us.agree ru.agree
##   <int>    <dbl>    <dbl>
## 1  1946    0.585    0.362
## 2  1947    0.621    0.383
## 3  1948    0.578    0.279
## 4  1949    0.541    0.377
## 5  1950    0.635    0.312
## 6  1951    0.487    0.402
```

Trends in global ideology

```
ggplot(data = annual.agree) +
  geom_line(mapping = aes(x = Year, y = us.agree), color = "blue") +
  geom_line(mapping = aes(x = Year, y = ru.agree), color = "red") +
  geom_text(aes(x = 2000, y = 0, label = "Voting with US"), color = "blue", data = data.frame()) +
  geom_text(aes(x = 2000, y = 1, label = "Voting with Russia"), color = "red", data = data.frame()) +
  geom_vline(aes(xintercept = 1989), linetype = "dotdash", color = "black") +
  geom_text(aes(x = 1993, y = 0.5, label = "Cold War Ends"), color = "black") +
  ylab("Proportion voting with Superpower") + theme_classic()
```

Grouping observations

Which side are you on?

Grouping countries: FP Similarity measures

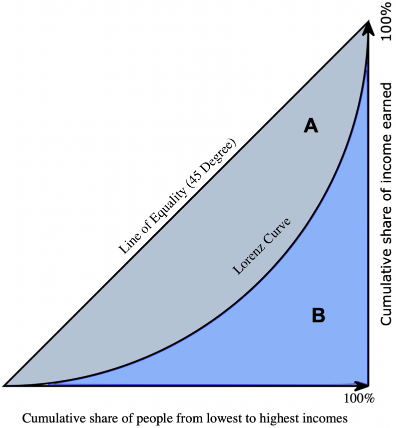
```
# Table for voting close to US
# USA
mydata %>%
  group_by(CountryName) %>%
  summarise(mean.pctUS = mean(PctAgreeUS)) %>%
  arrange(desc(mean.pctUS)) %>%
  head(n = 11) %>%
  filter(CountryName != "United States of America")
```

```
## # A tibble: 10 x 2
##    CountryName                     mean.pctUS
##    <chr>                                <dbl>
##  1 Palau                                0.736
##  2 United Kingdom                       0.652
##  3 Taiwan                               0.643
##  4 Israel                               0.640
##  5 Federated States of Micronesia      0.594
##  6 Canada                               0.586
##  7 Luxembourg                           0.571
##  8 Netherlands                          0.562
##  9 Belgium                              0.562
## 10 France                               0.549
```

Political polarization: QSS textbook

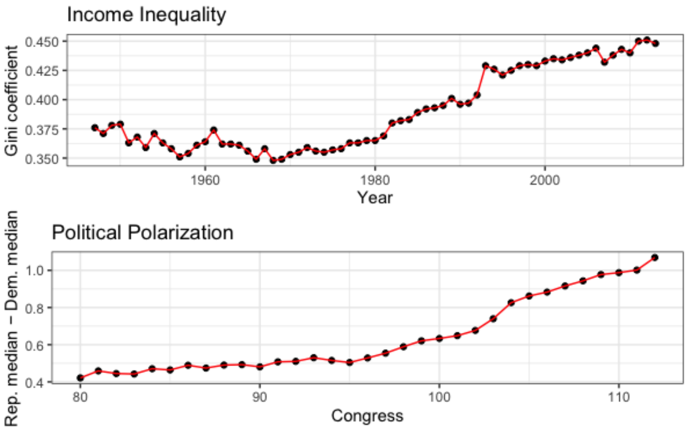
Income inequality $\rightarrow$ political polarization.

The *Gini coefficient*



US test case

Gini coefficient - Political Polarization



Association b-w variables

Income inequality $\rightarrow$ Political polarization?

Correlation:

- Summary of bivariate relationship.
- How two factors 'move together' on average.
- Always relative to mean value.

Product of z-scores:

$$cor(x, y) = \frac{1}{n}\sum_{i=1}^{n}(Z - x_i * Z - y_i)$$

Z-scores

A measure for the deviation from the mean (in SD terms)

Standardize variable

Allows comparison with *common units*

$$Zscore(X_i) = \frac{x_i - \bar{x}}{SD(X_i)}$$

Z score $> 0 \rightarrow$ unit larger than mean
Z score $< 0 \rightarrow$ unit smaller than mean

z-score example

HIGHEST-PAID ATHLETES IN THE WORLD

via Forbes' list of the highest-paid celebrities

		EARNINGS (JUNE 2019-MAY 2020)
1	Roger Federer	$106.3M
2	Cristiano Ronaldo	$105M
3	Lionel Messi	$104M
4	Neymar	$95.5M
5	LeBron James	$88.2M
6	Stephen Curry	$74.4M
7	Kevin Durant	$63.9M
8	Tiger Woods	$62.3M
9	Kirk Cousins	$60.5M
10	Carson Wentz	$59.1M

z-score example: QB salary

```
head(qb_data, n=10)
```

```
## # A tibble: 10 x 3
##    Player           Team     Avg_salary
##    <chr>            <chr>         <dbl>
##  1 Patrick Mahomes  Chiefs     45000000
##  2 Josh Allen       Bills      43005667
##  3 Dak Prescott     Cowboys    40000000
##  4 Deshaun Watson   Texans     39000000
##  5 Russell Wilson   Seahawks   35000000
##  6 Aaron Rodgers    Packers    33500000
##  7 Jared Goff       Lions      33500000
##  8 Kirk Cousins     Vikings    33000000
##  9 Carson Wentz     Colts      32000000
## 10 Matt Ryan        Falcons    30000000
```

z-score example: QB salary

```r
mean(qb_data$Avg_salary)
```

```
## [1] 33200378
```

```r
sd(qb_data$Avg_salary)
```

```
## [1] 6265045
```

```r
# Cousins z-score
((33000000 - mean(qb_data$Avg_salary))/sd(qb_data$Avg_salary))
```

```
## [1] -0.03198346
```

```r
# Mahomes z-score
((45000000 - mean(qb_data$Avg_salary))/sd(qb_data$Avg_salary))
```

```
## [1] 1.883406
```

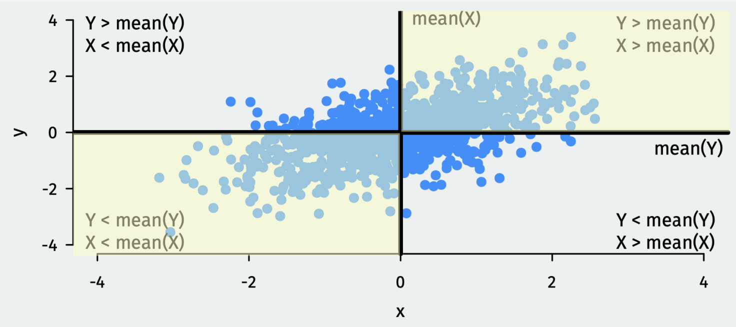
Outliers $\rightarrow$ more than 3 SD from mean

Correlation

- Average product of z-scores:
  - Positive correlation: when x is bigger than its mean, so is y
  - Negative correlation: when x is bigger than its mean, y is smaller
- z-score: not sensitive to unit used
- Correlation is identical even for different measuring units of variable

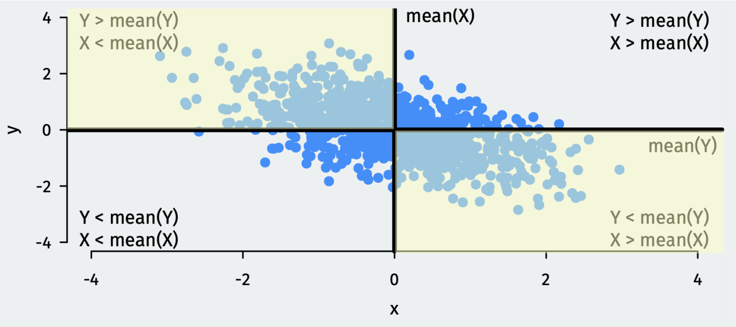
Correlation - how do the data look?

Positive correlation



Correlation - how do the data look?

Negative correlation



Correlation

- Measures **linear** association
- Order does not matter: cor(x,y) = cor(y,x)
- Interpretation:
  - Values range between (-1) to 1.
  - Close to 'edges' $\rightarrow$ stronger association.
  - Value of zero $\rightarrow$ no association.
  - Positive correlation $\rightarrow$ positive association.
  - Negative correlation $\rightarrow$ negative association.

Correlation in R

UN Voting: association b-w ideal point & liberal FP approach

```
# Voting with US
cor(mydata$idealpoint, mydata$PctAgreeUS, use = "pairwise")
```

```
## [1] 0.7498446
```

```
# Voting with Russia
cor(mydata$idealpoint, mydata$PctAgreeRUSSIA, use = "pairwise")
```

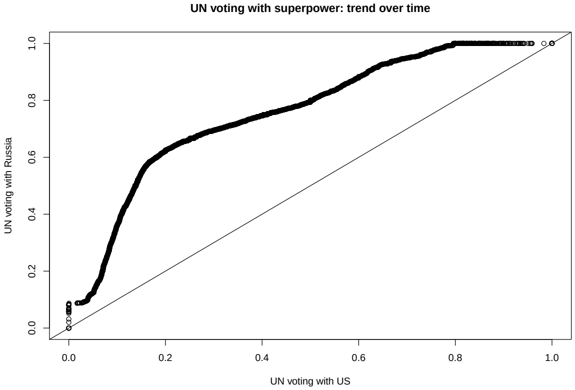
```
## [1] -0.7050107
```

Visualizing distributions

Qunatile Qunatile Plot

Scatter-plot of quantiles

```
### Q-Q plot
qqplot(mydata$PctAgreeUS, mydata$PctAgreeRUSSIA, xlab = "UN voting with US",
       ylab = "UN voting with Russia",
       main = "UN voting with superpower: trend over time")
abline(0,1)
```



Matrix in R

- Rectangular array with multiple values.
- Stores numeric variable (unlike data frame).
- Extract values with indexing [row, col].

```
### Build a matrix
m <- matrix(1:16, nrow = 4, ncol = 4, byrow = TRUE)
rownames(m) <- c("A","B","C","D")
colnames(m) <- c("W","X","Y","Z")
m
```

```
##    W  X  Y  Z
## A  1  2  3  4
## B  5  6  7  8
## C  9 10 11 12
## D 13 14 15 16
```

Working with matrices

Use math and apply functions

```
rowSums(m)
```

```
##  A  B  C  D
## 10 26 42 58
```

```
colMeans(m)
```

```
##  W  X  Y  Z
##  7  8  9 10
```

```
apply(m,1,mean)
```

```
##    A    B    C    D
##  2.5  6.5 10.5 14.5
```

```
apply(m,2,sd)
```

```
##        W        X        Y        Z
## 5.163978 5.163978 5.163978 5.163978
```

Lists in R

- General class of objects.
- Useful for storing multiple object types.

```
x <- list(y1 = c("this","is","a list", "of", "aggie", "games"),
          y2 = 1:5,
          y3 = data.frame(z1 = 1:4, z2 = c("Kent St.","Colorado","New Mexico"
                                            ,"Arkansas"),
                          z3 = c("Win","Win","Win","Loss")))

x$y3
```

```
##   z1         z2   z3
## 1  1   Kent St.  Win
## 2  2   Colorado  Win
## 3  3 New Mexico  Win
## 4  4   Arkansas Loss
```

```
x$y1
```

```
## [1] "this"    "is"      "a list" "of"      "aggie"   "games"
```

Clustering

- Identify associations within our data.
- Searching for *clusters* within large datasets.

- UN Voting data: diversity of global ideologies.
- Are there ‘clusters’ of ideologies?

Clustering

k-Means algorithm:

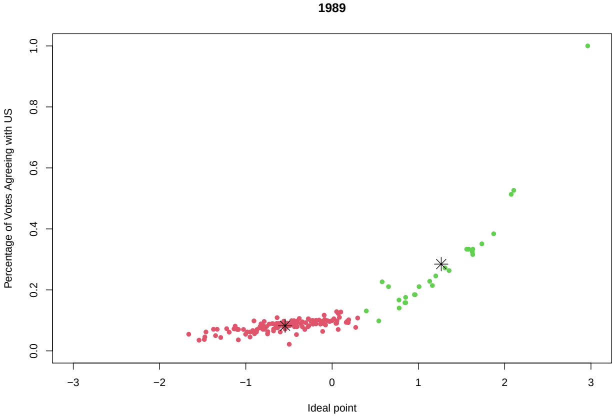
- *Iterative*: performed repeatedly to find differences b-w groups.
- Goal: split data to multiple similar groups (k-clusters).
- Each cluster is associated with a *centroid* (within group mean).

How?

- Observation assigned to closest cluster.
- Compute centroid based on new cluster.
- Researcher select initial number of clusters (k).
- Standardize data before procedure.

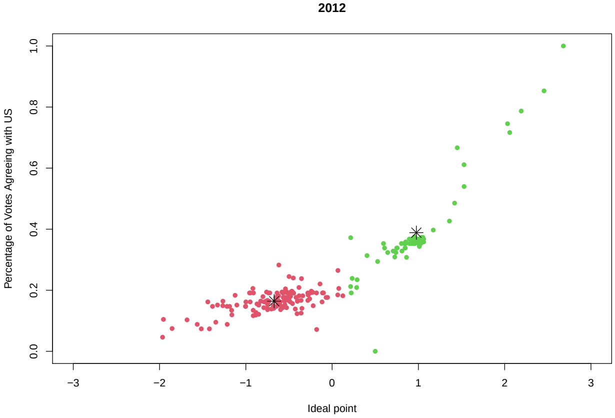
Cluster UN voting: 1989

```
# 1989 plot
un89 <- subset(mydata, subset = (Year == 1989))
cluster89 <- kmeans(un89[, c("idealpoint", "PctAgreeUS")], centers = 2)
un89$cluster1 <- cluster89$cluster

plot(x = un89$idealpoint, y = un89$PctAgreeUS, main = "1989",
     xlab = "Ideal point", ylab = "Percentage of Votes Agreeing with US",
     xlim = c(-3, 3), ylim = c(0, 1), pch = 16, col = un89$cluster1 + 1)
points(cluster89$centers, pch = 8, cex = 2) # add centroids
```



Cluster UN voting: 2012

```
## plot for 2012
un12 <- subset(mydata, subset = (Year == 2012))
cluster12 <- kmeans(un12[, c("idealpoint", "PctAgreeUS")], centers = 2)
un12$cluster2 <- cluster12$cluster

plot(x = un12$idealpoint, y = un12$PctAgreeUS, main = "2012",
     xlab = "Ideal point", ylab = "Percentage of Votes Agreeing with US",
     xlim = c(-3, 3), ylim = c(0, 1), pch = 16, col = un12$cluster2 + 1)
points(cluster12$centers, pch = 8, cex = 2)
```



UN data: shifting ideologies

Liberal $\rightarrow$ non-Liberal

```
## going from liberal cluster to non-liberal
un8912$CountryName[un8912$cluster1 > un8912$cluster2]
 [1] "Bahamas"                          "Cuba"                     "Haiti"
 [4] "Dominican Republic"               "Jamaica"                  "Trinidad and Tobago"
 [7] "Barbados"                         "Grenada"                  "St. Lucia"
[10] "St. Vincent and the Grenadines"   "Antigua & Barbuda"        "St. Kitts and Nevis"
[13] "Mexico"                           "Belize"                   "Guatemala"
[16] "Honduras"                         "El Salvador"              "Nicaragua"
[19] "Costa Rica"                       "Colombia"                 "Venezuela"
[22] "Guyana"                           "Suriname"                 "Ecuador"
[25] "Peru"                             "Brazil"                   "Bolivia"
[28] "Paraguay"                         "Argentina"                "Uruguay"
[31] NA                                 NA                         "Russia"
[34] "Belarus"                          "Cape Verde"               "Sao Tome and Principe"
[37] "Guinea-Bissau"                    "Equatorial Guinea"        "Gambia"
[40] "Mali"                             "Senegal"                  "Benin"
[43] "Mauritania"                       "Niger"                    "Ivory Coast"
[46] "Guinea"                           "Burkina Faso"             "Liberia"
[49] "Sierra Leone"                     "Ghana"                    "Togo"
```

UN data: shifting ideologies

non-Liberal $\rightarrow$ Liberal

```
> ## going from non-liberal to liberal cluster
> un8912$CountryName[un8912$cluster1 < un8912$cluster2]
 [1] "United States of America" "Canada"                   "United Kingdom"           "Ireland"                  "Netherlands"
 [6] "Belgium"                  "Luxembourg"               "France"                   "Spain"                    "Portugal"
[11] "German Federal Republic"  NA                         "Austria"                  NA                         "Italy"
[16] "Malta"                    "Greece"                   "Finland"                  "Sweden"                   "Norway"
[21] "Denmark"                  "Iceland"                  "Turkey"                   "Israel"                   NA
[26] "Japan"                    "Australia"                "New Zealand"
>
```

Wrapping up week 5

Summary:

- Measuring complex (latent) concepts: terrorism, resolve.
- Visualize bivariate relations: scatter plot.
- z-scores and standardizing units.
- Correlation: how two factors 'move together'.
- Clustering: explore similarities in large dataset.
- R work: scatterplots, cor(), qqplot(), matrix(), list(), kmean()

Task 2: Working with R:

- Canvas (Wed/Thu.), more details next week.