The Application of 2-D Visualization on Multi-Dimensional Data

劉自平（Tzu-Ping Liu）[1]
tpliu@utaipei.edu.tw

[1]台北市立大學社會暨公共事務系

Outline

- Why Data Visualization
- Scaling—Common Methods Within Political Science
- Newly Emerging Approach—Contrastive Learning
- Comparison between the Two
- Discussion

Why Visualization

- To present the information of data graphically:

Why Visualization

- To present the information of data graphically:
  - Straightforward

Why Visualization

- To present the information of data graphically:
  - Straightforward
  - Intuitive

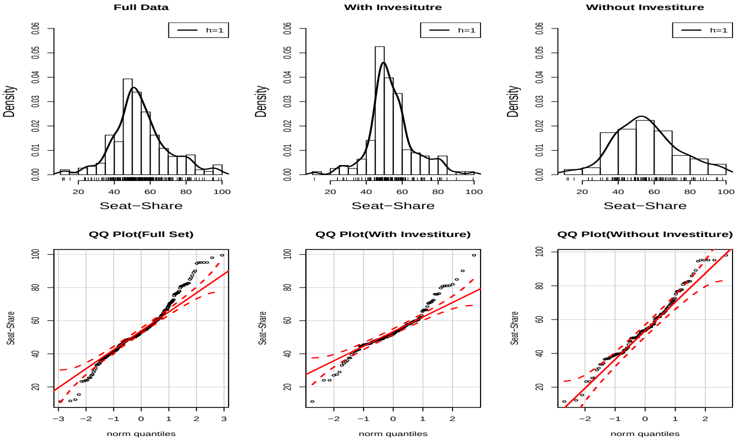
Why Visualization

- To present the information of data graphically:
  - Straightforward
  - Intuitive
  - Simple

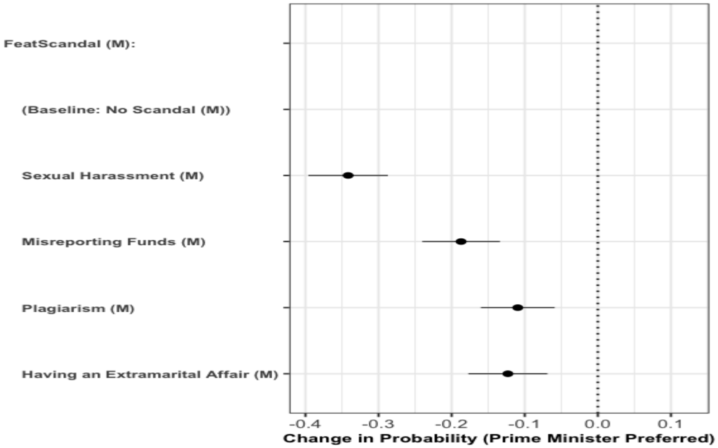
Why Visualization

- To present the information of data graphically:
  - Straightforward
  - Intuitive
  - Simple
  - Parsimonious

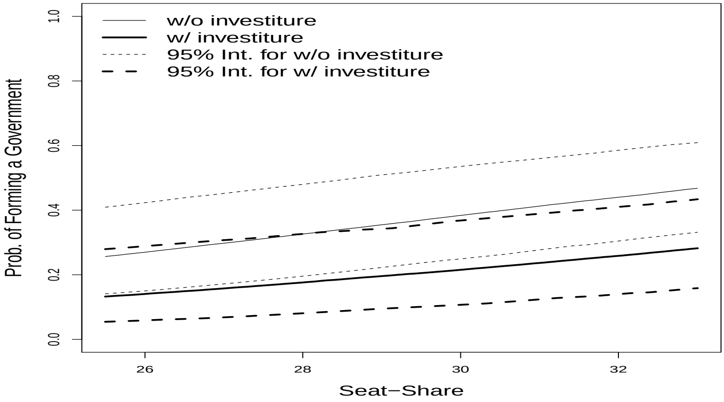
Example—Visualizing Validation of Assumption



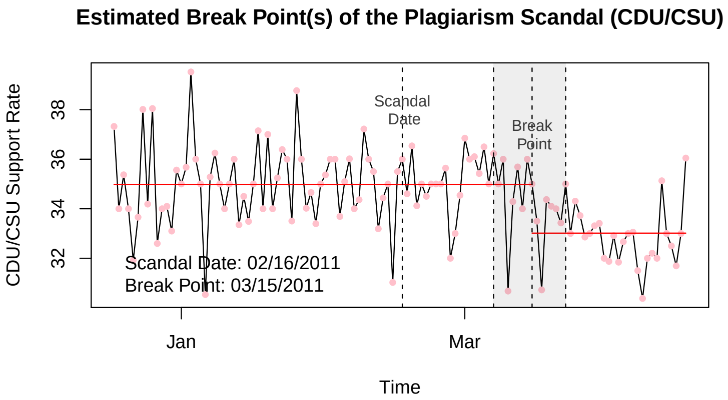
Example—Visualizing Regression Estimation



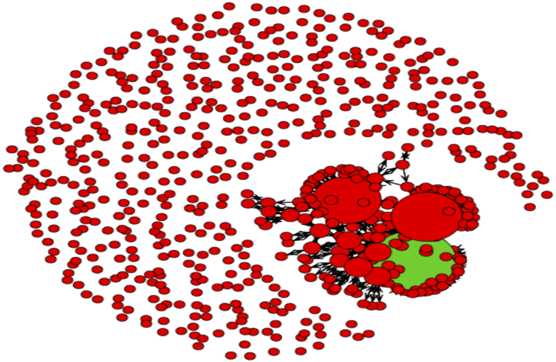
Example—Visualizing Prediction



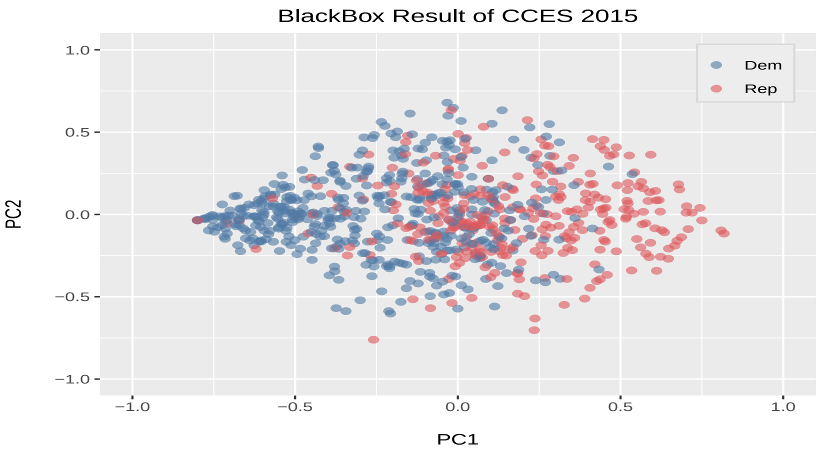
Example—Visualizing Trend



Example—Visualizing Relationships



Example—Visualizing Latent Traits (Today's Focus)



What is Scaling

- (Mathematical) Definition of scaling:

What is Scaling

- (Mathematical) Definition of scaling:
  - Change the size (mainly) and the structure (slightly) to approximate an object, e.g., image, data, etc.

What is Scaling

- (Mathematical) Definition of scaling:
  - Change the size (mainly) and the structure (slightly) to approximate an object, e.g., image, data, etc.
- It sounds like regression-type models:

What is Scaling

- (Mathematical) Definition of scaling:
  - Change the size (mainly) and the structure (slightly) to approximate an object, e.g., image, data, etc.
- It sounds like regression-type models:
  - In regression, what do we approximate about?

What is Scaling

- (Mathematical) Definition of scaling:
  - Change the size (mainly) and the structure (slightly) to approximate an object, e.g., image, data, etc.
- It sounds like regression-type models:
  - In regression, what do we approximate about?
  - The regression line

What is Scaling

- (Mathematical) Definition of scaling:
  - Change the size (mainly) and the structure (slightly) to approximate an object, e.g., image, data, etc.
- It sounds like regression-type models:
  - In regression, what do we approximate about?
  - The regression line
- What is the main difference between the two?

What is Scaling

- (Mathematical) Definition of scaling:
  - Change the size (mainly) and the structure (slightly) to approximate an object, e.g., image, data, etc.
- It sounds like regression-type models:
  - In regression, what do we approximate about?
  - The regression line
- What is the main difference between the two?
  - Regression: Estimating the relationships between X and Y (Supervised Learning)

What is Scaling

- (Mathematical) Definition of scaling:
  - Change the size (mainly) and the structure (slightly) to approximate an object, e.g., image, data, etc.
- It sounds like regression-type models:
  - In regression, what do we approximate about?
  - The regression line
- What is the main difference between the two?
  - Regression: Estimating the relationships between X and Y (Supervised Learning)
  - Scaling: Through X to define unknown Y (Unsupervised Learning)

What is Scaling For

- Purposes for scaling methods:

What is Scaling For

- Purposes for scaling methods:
  - Data reduction: parsimony and the curse of dimensionality

What is Scaling For

- Purposes for scaling methods:
  - Data reduction: parsimony and the curse of dimensionality
  - Testing dimensionality: what are the sources of variability between objects?

What is Scaling For

- Purposes for scaling methods:
  - Data reduction: parsimony and the curse of dimensionality
  - Testing dimensionality: what are the sources of variability between objects?
  - Measuring objects: need measures of concepts of interest

What is Scaling For

- Purposes for scaling methods:
  - Data reduction: parsimony and the curse of dimensionality
  - Testing dimensionality: what are the sources of variability between objects?
  - Measuring objects: need measures of concepts of interest
  - Graphical display: numbers vs. graphics

Scaling and Spatial Theory

- Spatial Voting:

Scaling and Spatial Theory

- Spatial Voting:
  - Everyone's political preference can be represented by their positions on an ideological scale, and citizens' prefer candidates' whose positions are closer to their own

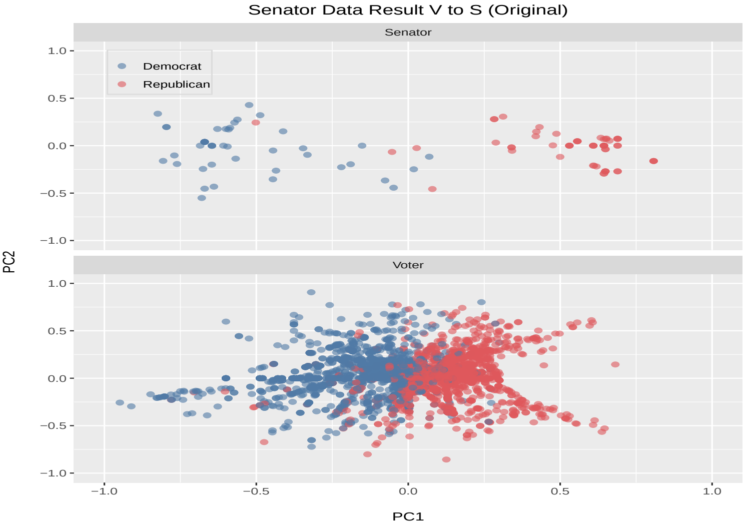
Scaling and Spatial Theory

- Spatial Voting:
  - Everyone's political preference can be represented by their positions on an ideological scale, and citizens' prefer candidates' whose positions are closer to their own
- Scaling methods:

Scaling and Spatial Theory

- Spatial Voting:
  - Everyone's political preference can be represented by their positions on an ideological scale, and citizens' prefer candidates' whose positions are closer to their own
- Scaling methods:
  - Produce a geometric representation of latent trait within the data

Example of Scaling Results



Common Scaling Methods

- There are several popular methods in general and in political science:

Common Scaling Methods

- There are several popular methods in general and in political science:
  - Nonparametric or semi-parametric:

Common Scaling Methods

- There are several popular methods in general and in political science:
    - Nonparametric or semi-parametric:
        - Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA)

Common Scaling Methods

- There are several popular methods in general and in political science:
  - Nonparametric or semi-parametric:
    - Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA)
    - Optimal Classification (OC) and Ordered Optimal Classification (OOC)

Common Scaling Methods

- There are several popular methods in general and in political science:
  - Nonparametric or semi-parametric:
    - Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA)
    - Optimal Classification (OC) and Ordered Optimal Classification (OOC)
  - Parametric:

Common Scaling Methods

- There are several popular methods in general and in political science:
  - Nonparametric or semi-parametric:
    - Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA)
    - Optimal Classification (OC) and Ordered Optimal Classification (OOC)
  - Parametric:
    - Factor Analysis

Common Scaling Methods

- There are several popular methods in general and in political science:
  - Nonparametric or semi-parametric:
    - Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA)
    - Optimal Classification (OC) and Ordered Optimal Classification (OOC)
  - Parametric:
    - Factor Analysis
    - Aldrich-McKelvey Scaling

Common Scaling Methods

- There are several popular methods in general and in political science:
  - Nonparametric or semi-parametric:
    - Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA)
    - Optimal Classification (OC) and Ordered Optimal Classification (OOC)
  - Parametric:
    - Factor Analysis
    - Aldrich-McKelvey Scaling
    - Black-box scaling

Common Scaling Methods

- There are several popular methods in general and in political science:
  - Nonparametric or semi-parametric:
    - Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA)
    - Optimal Classification (OC) and Ordered Optimal Classification (OOC)
  - Parametric:
    - Factor Analysis
    - Aldrich-McKelvey Scaling
    - Black-box scaling
    - Item Response Theory (IRT)

Common Scaling Methods

- There are several popular methods in general and in political science:
  - Nonparametric or semi-parametric:
    - Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA)
    - Optimal Classification (OC) and Ordered Optimal Classification (OOC)
  - Parametric:
    - Factor Analysis
    - Aldrich-McKelvey Scaling
    - Black-box scaling
    - Item Response Theory (IRT)
    - NOMINATE Scores

Principal Component Analysis

- Main ideas behind PCA:

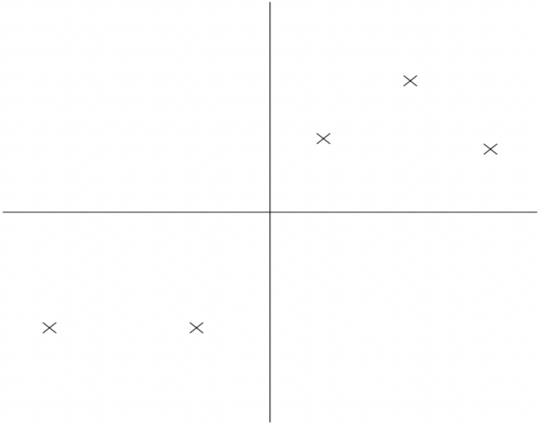
Principal Component Analysis

- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)

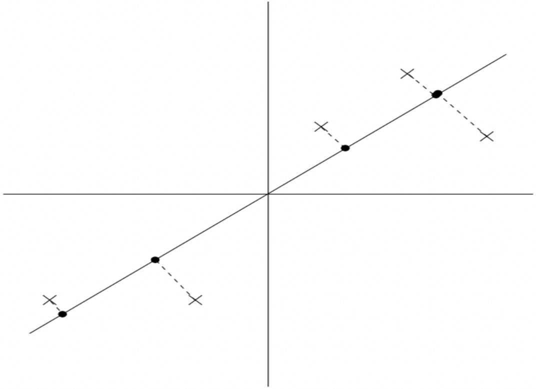
Principal Component Analysis

- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)
  - The dimensions/directions are those on which data points distribute/vary maximally

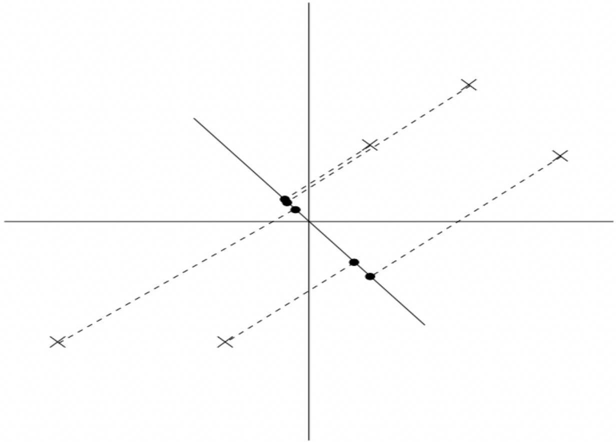
PCA Demonstration I (Andrew Ng, 2015)



PCA Demonstration II (Andrew Ng, 2015)



PCA Demonstration III (Andrew Ng, 2015)



Principal Component Analysis

- Main ideas behind PCA:

Principal Component Analysis

- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)
  - The dimensions/directions are those on which data points distribute/vary maximally
- PCA can be used for:

Principal Component Analysis

- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)
  - The dimensions/directions are those on which data points distribute/vary maximally
- PCA can be used for:
  - Dimension reduction: reducing the amount of variables (multi-collinearity)

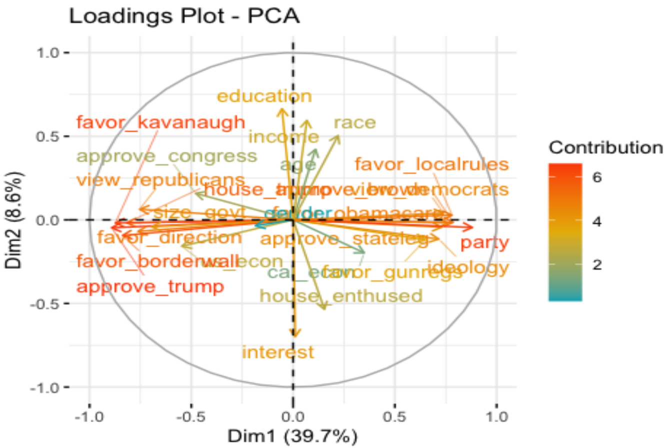
Principal Component Analysis

- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)
  - The dimensions/directions are those on which data points distribute/vary maximally
- PCA can be used for:
  - Dimension reduction: reducing the amount of variables (multi-collinearity)
  - Measurement: Measuring concepts which are combined by variables, e.g., ideology, ability, etc.

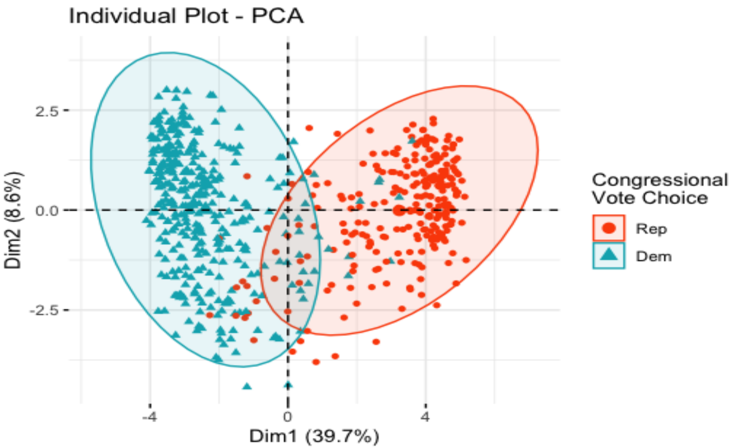
Principal Component Analysis

- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)
  - The dimensions/directions are those on which data points distribute/vary maximally
- PCA can be used for:
  - Dimension reduction: reducing the amount of variables (multi-collinearity)
  - Measurement: Measuring concepts which are combined by variables, e.g., ideology, ability, etc.
  - Visualization: Intuitive interpretation

Application of PCA—Loading Plot



Application of PCA—Individual Plot



Principal Component Analysis

- Main ideas behind PCA:

Principal Component Analysis

- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)
  - The dimensions/directions are those on which data points distribute/vary maximally
- PCA can be used for:

Principal Component Analysis

- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)
  - The dimensions/directions are those on which data points distribute/vary maximally
- PCA can be used for:
  - Dimension reduction: reducing the amount of variables (multi-collinearity)
  - Measurement: Measuring concepts which are combined by variables, e.g., ideology, ability, etc.
  - Visualization: Intuitive interpretation
- Issues:

Principal Component Analysis

- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)
  - The dimensions/directions are those on which data points distribute/vary maximally
- PCA can be used for:
  - Dimension reduction: reducing the amount of variables (multi-collinearity)
  - Measurement: Measuring concepts which are combined by variables, e.g., ideology, ability, etc.
  - Visualization: Intuitive interpretation
- Issues:
  - Continuous and ordinal data only (is ordinal data truly ordinal?)

Principal Component Analysis

- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)
  - The dimensions/directions are those on which data points distribute/vary maximally
- PCA can be used for:
  - Dimension reduction: reducing the amount of variables (multi-collinearity)
  - Measurement: Measuring concepts which are combined by variables, e.g., ideology, ability, etc.
  - Visualization: Intuitive interpretation
- Issues:
  - Continuous and ordinal data only (is ordinal data truly ordinal?)
  - Missing data unavailable

Principal Component Analysis

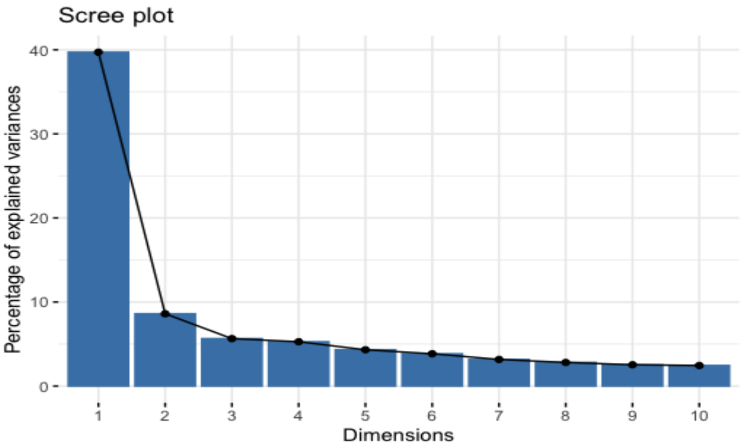
- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)
  - The dimensions/directions are those on which data points distribute/vary maximally
- PCA can be used for:
  - Dimension reduction: reducing the amount of variables (multi-collinearity)
  - Measurement: Measuring concepts which are combined by variables, e.g., ideology, ability, etc.
  - Visualization: Intuitive interpretation
- Issues:
  - Continuous and ordinal data only (is ordinal data truly ordinal?)
  - Missing data unavailable
- How to derive principal components:

Principal Component Analysis

- Main ideas behind PCA:
  - Transferring a multi-dimensional data into lower dimensional representation (usually two dimensions, why?)
  - The dimensions/directions are those on which data points distribute/vary maximally
- PCA can be used for:
  - Dimension reduction: reducing the amount of variables (multi-collinearity)
  - Measurement: Measuring concepts which are combined by variables, e.g., ideology, ability, etc.
  - Visualization: Intuitive interpretation
- Issues:
  - Continuous and ordinal data only (is ordinal data truly ordinal?)
  - Missing data unavailable
- How to derive principal components:
  - Decompose the variance-covariance matrix of data to find eigenvectors

How Many PCs?

- The elbow method



Multiple Correspondence Analysis

- MCA is invented to deal with PCA's issues

Multiple Correspondence Analysis

- MCA is invented to deal with PCA's issues
  - Binary, ordinal, and nominal data types are all available

Multiple Correspondence Analysis

- MCA is invented to deal with PCA's issues
  - Binary, ordinal, and nominal data types are all available
  - Continuous data may not be excluded from the analysis necessarily

Multiple Correspondence Analysis

- MCA is invented to deal with PCA's issues
  - Binary, ordinal, and nominal data types are all available
  - Continuous data may not be excluded from the analysis necessarily
  - No missing data issue
- Ideas behind MCA:
  - Treating each value as a dummy

Multiple Correspondence Analysis

- MCA is invented to deal with PCA's issues
  - Binary, ordinal, and nominal data types are all available
  - Continuous data may not be excluded from the analysis necessarily
  - No missing data issue
- Ideas behind MCA:
  - Treating each value as a dummy
  - Transferring this binary data into probability data (as if continuous data)

Multiple Correspondence Analysis

- MCA is invented to deal with PCA's issues
  - Binary, ordinal, and nominal data types are all available
  - Continuous data may not be excluded from the analysis necessarily
  - No missing data issue
- Ideas behind MCA:
  - Treating each value as a dummy
  - Transferring this binary data into probability data (as if continuous data)
  - Apply PCA process to the transferred data

Multiple Correspondence Analysis

- MCA is invented to deal with PCA's issues
  - Binary, ordinal, and nominal data types are all available
  - Continuous data may not be excluded from the analysis necessarily
  - No missing data issue
- Ideas behind MCA:
  - Treating each value as a dummy
  - Transferring this binary data into probability data (as if continuous data)
  - Apply PCA process to the transferred data
- Issue:

Multiple Correspondence Analysis

- MCA is invented to deal with PCA's issues
  - Binary, ordinal, and nominal data types are all available
  - Continuous data may not be excluded from the analysis necessarily
  - No missing data issue
- Ideas behind MCA:
  - Treating each value as a dummy
  - Transferring this binary data into probability data (as if continuous data)
  - Apply PCA process to the transferred data
- Issue:
  - The result may be too complicated to be formative

What is Contrastive Learning

- Analyzing high-dimensional data to capture “patterns that are specific to, or enriched in, one dataset relative to another”

What is Contrastive Learning

- Analyzing high-dimensional data to capture "patterns that are specific to, or enriched in, one dataset relative to another"
- Idea behind contrastive learning:

What is Contrastive Learning

- Analyzing high-dimensional data to capture “patterns that are specific to, or enriched in, one dataset relative to another”
- Idea behind contrastive learning:
  - Exploring unique components or directions that contain more salient latent patterns in one dataset (the target group) than the other (background group)

What is Contrastive Learning

- Analyzing high-dimensional data to capture “patterns that are specific to, or enriched in, one dataset relative to another”
- Idea behind contrastive learning:
  - Exploring unique components or directions that contain more salient latent patterns in one dataset (the target group) than the other (background group)
- Difference between ordinary approach and contrastive learning:

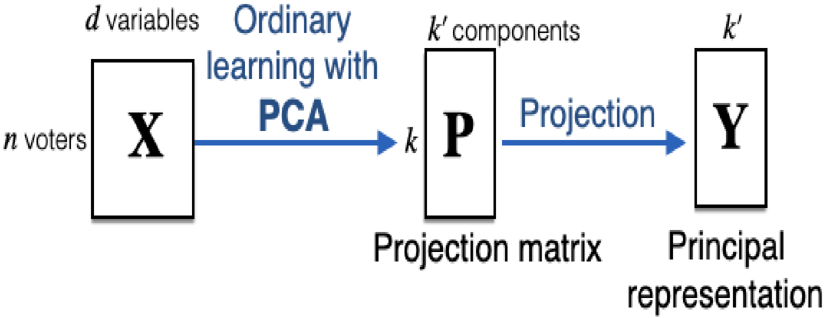
What is Contrastive Learning

- Analyzing high-dimensional data to capture “patterns that are specific to, or enriched in, one dataset relative to another”
- Idea behind contrastive learning:
  - Exploring unique components or directions that contain more salient latent patterns in one dataset (the target group) than the other (background group)
- Difference between ordinary approach and contrastive learning:
  - Ordinary scaling: Identifying principal components/directions on which the data *as a whole* varies maximally or distributes based on respondents’ (dis)similar responses to questions/stimuli

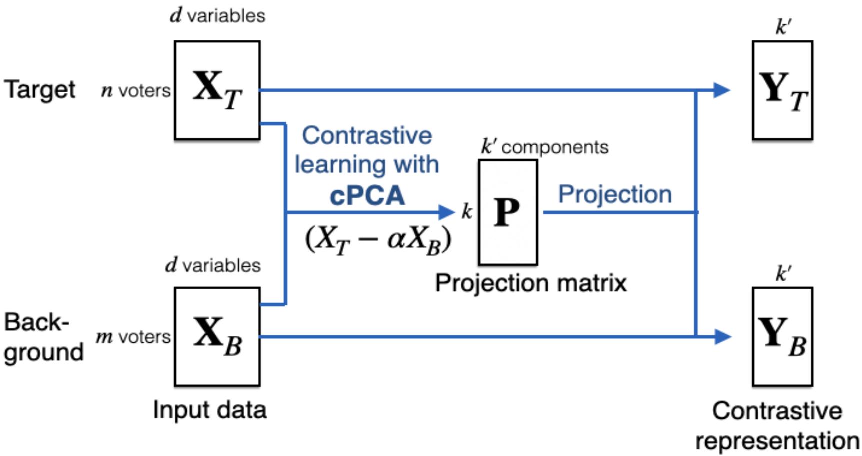
What is Contrastive Learning

- Analyzing high-dimensional data to capture "patterns that are specific to, or enriched in, one dataset relative to another"
- Idea behind contrastive learning:
  - Exploring unique components or directions that contain more salient latent patterns in one dataset (the target group) than the other (background group)
- Difference between ordinary approach and contrastive learning:
  - Ordinary scaling: Identifying principal components/directions on which the data *as a whole* varies maximally or distributes based on respondents' (dis)similar responses to questions/stimuli
  - Contrastive scaling: Splitting data into different groups first, usually by predefined classes (e.g., party ID), and then compares the data structure of the target group against the background group to find PC(s) on which the target group varies maximally and the background group varies minimally

Ordinary PCA Work Flow



Contrastive PCA Work Flow



Contrast Parameter—α

- While conducting contrastive learning, there is a contrast parameter—α:

Contrast Parameter—α

- While conducting contrastive learning, there is a contrast parameter—α:
  - α is a hyper parameter, i.e., researchers can tune α based on subjective or objective criteria

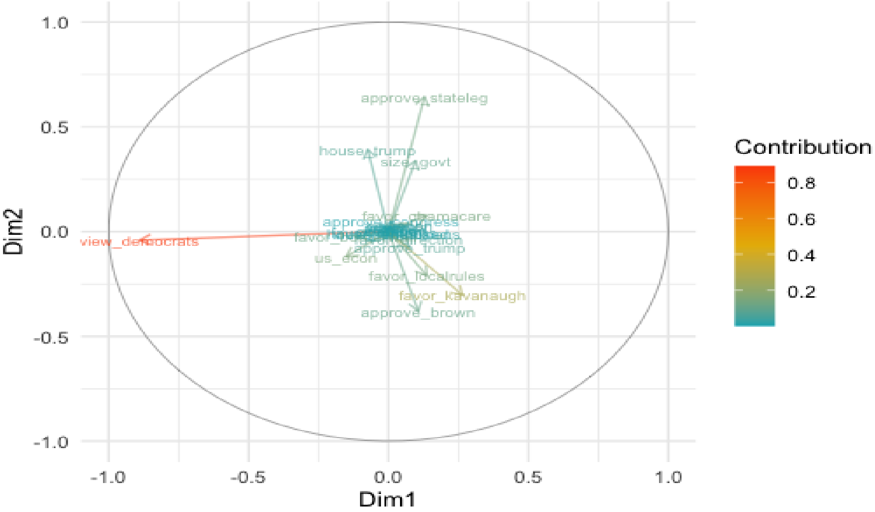
Contrast Parameter—α

- While conducting contrastive learning, there is a contrast parameter—α:
  - α is a hyper parameter, i.e., researchers can tune α based on subjective or objective criteria
  - One way to intuitively think about α is to treat it as the weight we want to put on the background group—while α is equal to 0, the procedure becomes to apply ordinary scaling on the target group

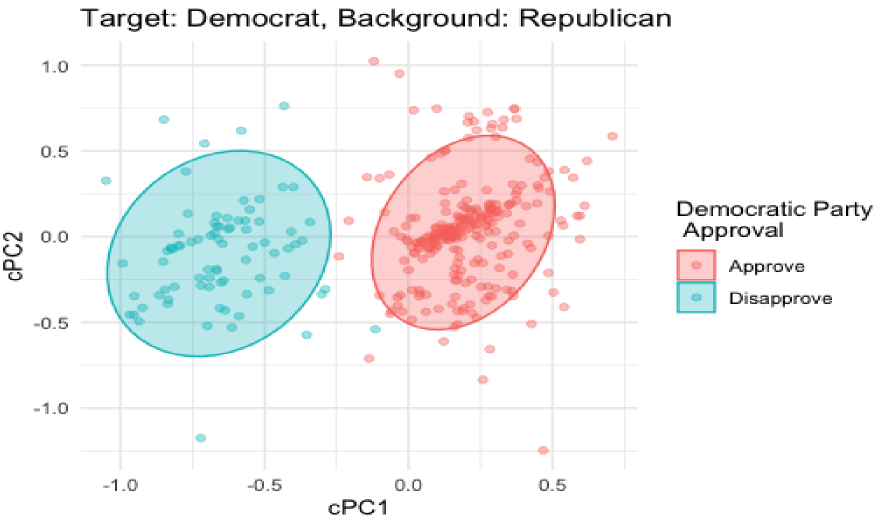
Contrast Parameter—α

- While conducting contrastive learning, there is a contrast parameter—α:
  - α is a hyper parameter, i.e., researchers can tune α based on subjective or objective criteria
  - One way to intuitively think about α is to treat it as the weight we want to put on the background group—while α is equal to 0, the procedure becomes to apply ordinary scaling on the target group
  - The other way to intuitive think about α is to treat α as a value set—different α means we would like to compare two groups through a different perspective

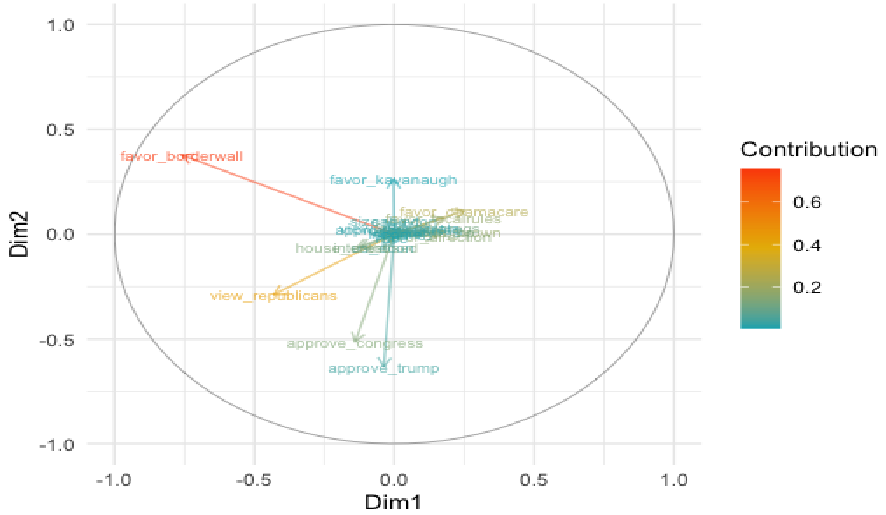
cPCA Loadings of Californian Voters (TG: Democrats)



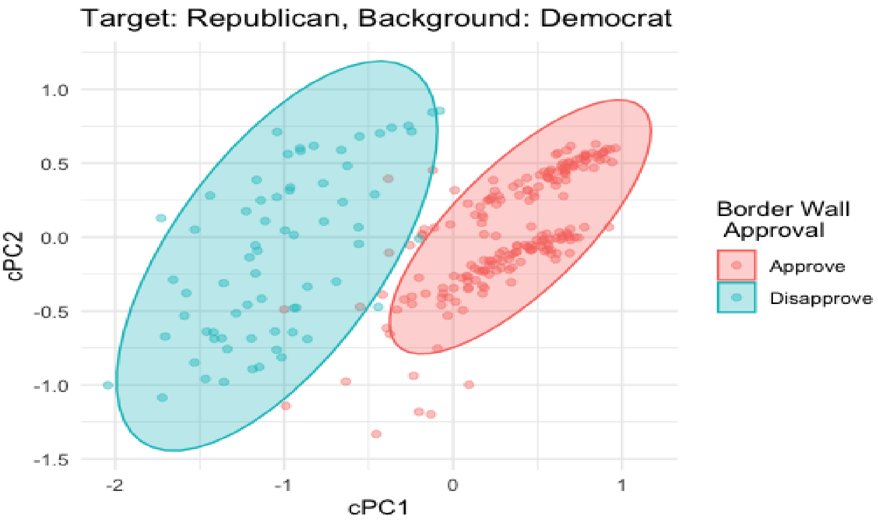
cPCA Individuals of Californian Voters (TG: Democrats)



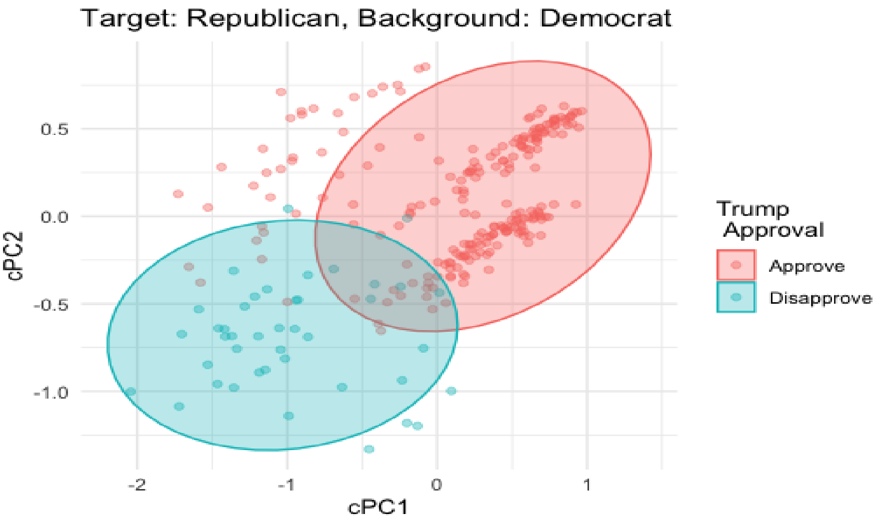
cPCA Loadings of Californian Voters (TG: Republicans)



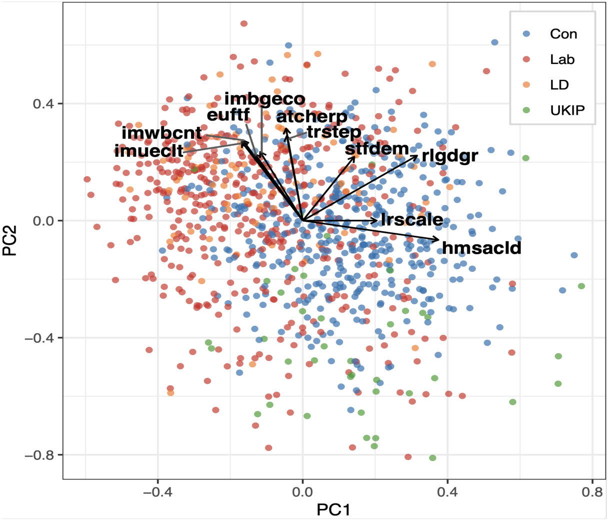
cPCA Individuals of Californian Voters—Wall (TG: Republicans)



cPCA Individuals of Californian Voters—Trump (TG: Republicans)

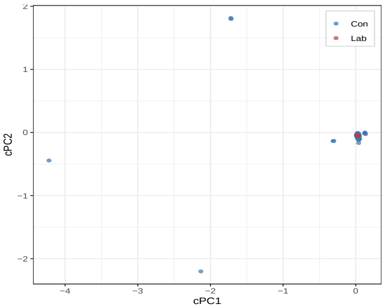


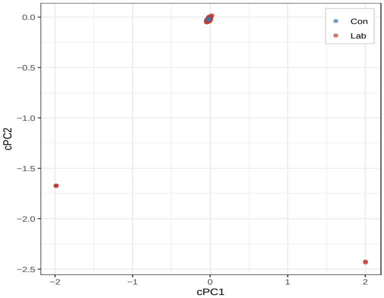
Ordinary Scaling (Black-Box Scaling) Results of ESS 2018 (UK Module)



atcherp: Emotionally Attached to Europe
euftf: European Unification
hmsacld: Gay Adoption
imbgeco: Immigration to Economy
imueclt: Immigration to Culture
imwbcnt: Immigration to Living Quality
lrscale: Left-Right Scale
stfdem: Democracy Satisfaction
trstep: Trust in the European Parliament
(Please refer to Appendix B for details)

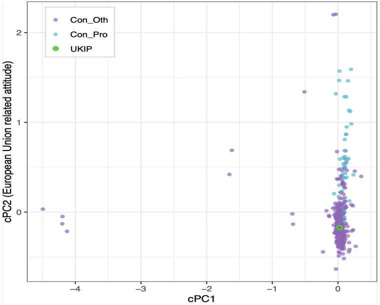
Contrastive MCA Results of ESS 2018 I (UK Module)

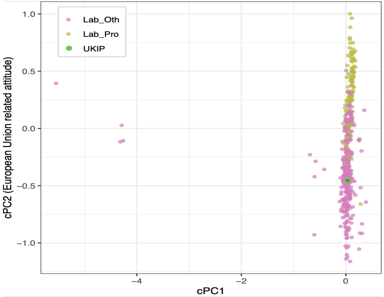




- TG: Conservatives, BG: Labours (Left Panel)
- TG: Labours, BG: Conservatives (Right Panel)

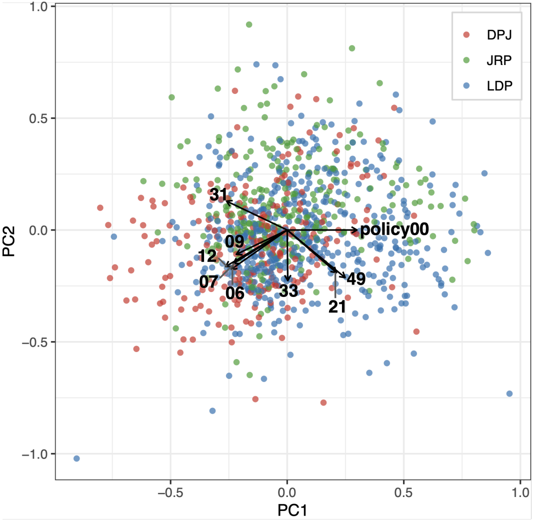
Contrastive MCA Results of ESS 2018 II (UK Module)





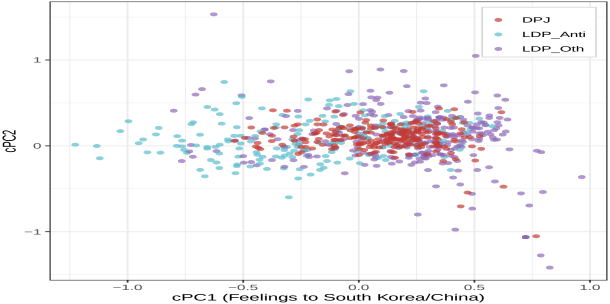
- TG: Conservatives, BG: UKIP (Left Panel)
- TG: Labours, BG: UKIP (Right Panel)

Ordinary Scaling (Black-Box Scaling) Results of UTAS 2012



policy00: Ideology
policy06: Constitution Change
policy07: Defense Capacity
policy09: Preemptive Attack
policy12: Collective Defense
policy21: Permanent Foreign Residents with Local Suffrage
policy31: Nuclear Power Plant Restart
policy33: Prime Minister Selection
policy49: Nuclear Power 2030
(Please refer to Appendix B for details)

Contrastive MCA Results of UTAS 2012



- TG: LDP, BG: DPJ

Scaling in General

- Scaling method can be utilized for:

Scaling in General

- Scaling method can be utilized for:
  - Preliminarily investigating latent traits buried in the data

Scaling in General

- Scaling method can be utilized for:
  - Preliminarily investigating latent traits buried in the data
  - Reducing dimensionality of data to prevent multicollinearity and inefficiency

Scaling in General

- Scaling method can be utilized for:
  - Preliminarily investigating latent traits buried in the data
  - Reducing dimensionality of data to prevent multicollinearity and inefficiency
  - Creating measurement for complex concepts

Scaling in General

- Scaling method can be utilized for:
  - Preliminarily investigating latent traits buried in the data
  - Reducing dimensionality of data to prevent multicollinearity and inefficiency
  - Creating measurement for complex concepts
- There is one issue existing in almost all current scaling methods:

Scaling in General

- Scaling method can be utilized for:
  - Preliminarily investigating latent traits buried in the data
  - Reducing dimensionality of data to prevent multicollinearity and inefficiency
  - Creating measurement for complex concepts
- There is one issue existing in almost all current scaling methods:
  - Almost all scaling methods only model the level of “likeness”

Scaling in General

- Scaling method can be utilized for:
  - Preliminarily investigating latent traits buried in the data
  - Reducing dimensionality of data to prevent multicollinearity and inefficiency
  - Creating measurement for complex concepts
- There is one issue existing in almost all current scaling methods:
  - Almost all scaling methods only model the level of “likeness”
  - In practice, these methods will consider “the Squad” to be closer to moderate Republicans

Scaling in General

- Scaling method can be utilized for:
  - Preliminarily investigating latent traits buried in the data
  - Reducing dimensionality of data to prevent multicollinearity and inefficiency
  - Creating measurement for complex concepts
- There is one issue existing in almost all current scaling methods:
  - Almost all scaling methods only model the level of "likeness"
  - In practice, these methods will consider "the Squad" to be closer to moderate Republicans
  - Duck-Mayr and Montegomery (2021) include the level of "dislikeness" in their scaling model which solves the aforementioned issue

Contrastive Scaling in General

- Utility of contrastive learning:

Contrastive Scaling in General

- Utility of contrastive learning:
  - Mining data from different perspectives which are overlooked by ordinary approaches

Contrastive Scaling in General

- Utility of contrastive learning:
  - Mining data from different perspectives which are overlooked by ordinary approaches
  - Checking similarity between datasets, e.g., balance-checking between the treatment and control groups

Contrastive Scaling in General

- Utility of contrastive learning:
  - Mining data from different perspectives which are overlooked by ordinary approaches
  - Checking similarity between datasets, e.g., balance-checking between the treatment and control groups
  - Providing objective insights for subgroup analysis

Causality

- Note that estimating or deriving causality is different:

Causality

- Note that estimating or deriving causality is different:
  - The latent traits within data only demonstrate correlation/association

Causality

- Note that estimating or deriving causality is different:
  - The latent traits within data only demonstrate correlation/association
  - Causal inference is mainly a design issue but not a data or modeling issue

Causality

- Note that estimating or deriving causality is different:
  - The latent traits within data only demonstrate correlation/association
  - Causal inference is mainly a design issue but not a data or modeling issue
  - Big data does not help much with causal inference

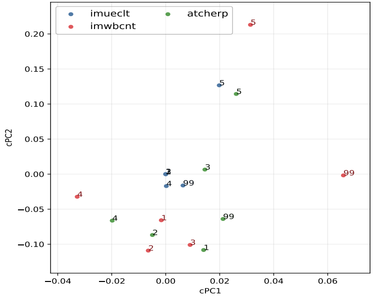
Causality

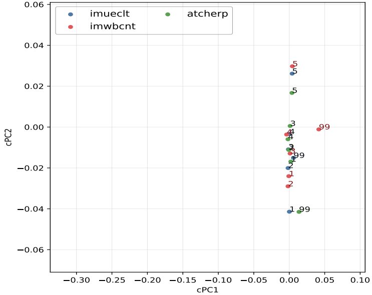
- Note that estimating or deriving causality is different:
  - The latent traits within data only demonstrate correlation/association
  - Causal inference is mainly a design issue but not a data or modeling issue
  - Big data does not help much with causal inference
  - Statistical model or machine learning may help only when the design is correct

The End

Thank You!

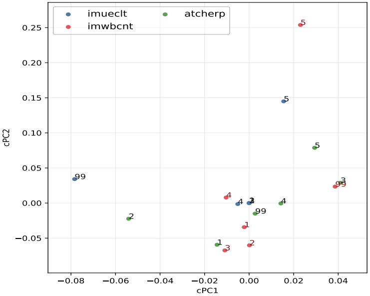
Loadings and Variable Coordinates I

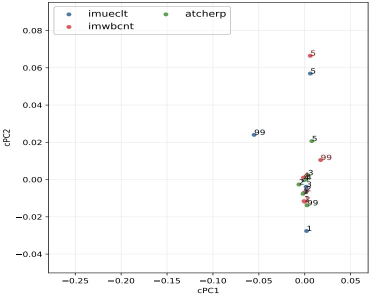




- Loadings for ESS 2018—TG: Labours, BG: UKIP (Left Panel)
- Variable Coordinates for ESS 2018—TG: Labours, BG: UKIP (Right Panel)

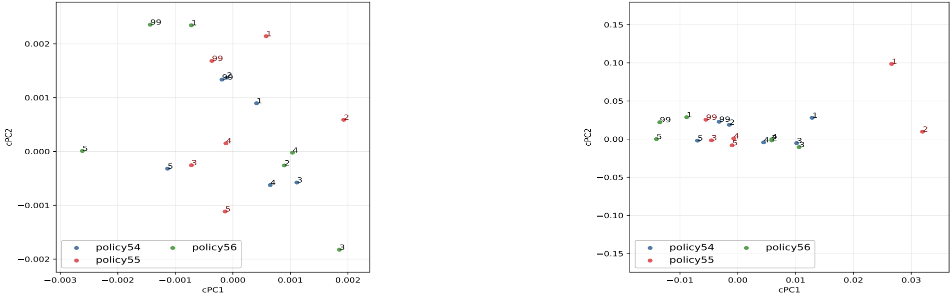
Loadings and Variable Coordinates II





- Loadings for ESS 2018—TG: Conservatives, BG: UKIP (Left Panel)
- Variable Coordinates for ESS 2018—TG: Conservatives, BG: UKIP (Right Panel)

Loadings and Variable Coordinates III



- Loadings for UTAS 2012—TG: LDP, BG: DPJ (Left Panel)
- Variable Coordinates for UTAS 2012—TG: LDP, BG: DPJ (Right Panel)