



Statistical Overview and Visualization of Hong Kong Legislative Council Election

Mentor: Dr. Kwai Wong

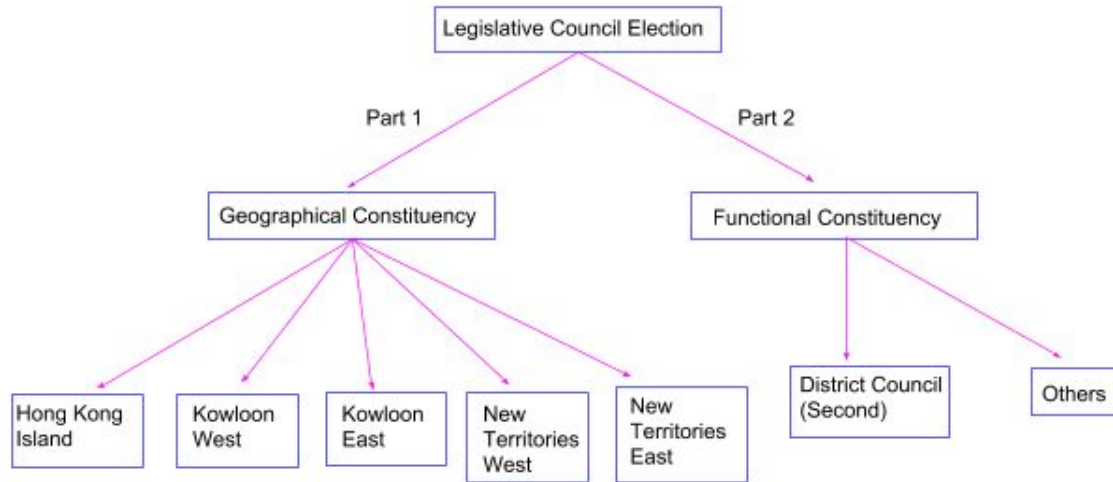
Project by:

NG Ching Tao

Jorge Garcia

Frank Betancourt

Hong Kong Legislative Council Election



- Part 2 Other Functional Constituency is not included
- Two parties of candidate:
 - Pro- government camp
 - Pro- choice camp

R programming

E.g.

```
cor.test(RP2012_sorted$q7, RP2012_sorted$R3)
```

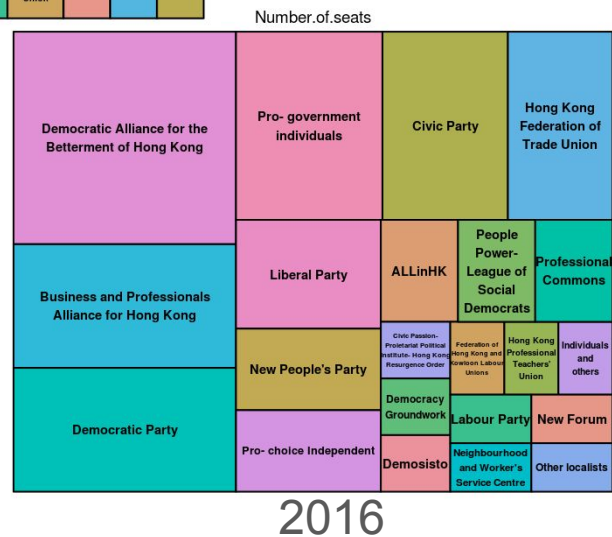
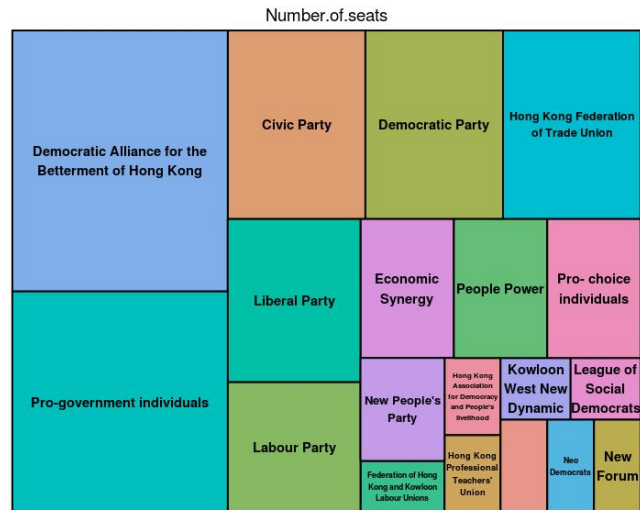
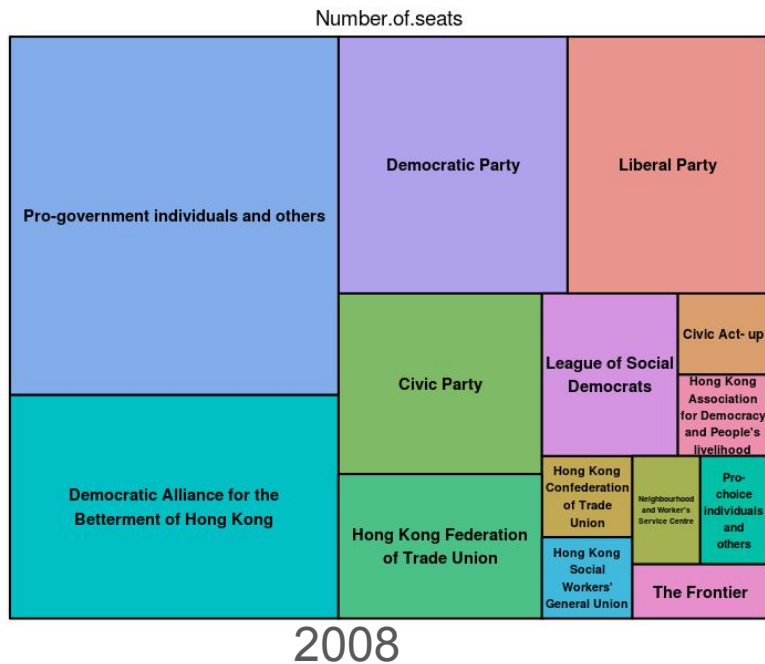
Pearson's product-moment correlation

sample estimates:

cor

0.06245718

Election Result



Data Correlation

In election period, the dominant factors are

	Hong Kong Island	Kowloon West	Kowloon East	New Territories West	New Territories East	District Council (Second)
2008	Political Inclination	Political Inclination	Political Inclination	Political Inclination	Political Inclination	/
2012	Emphasis on relationship with Central government raised by candidate	Follow strategic plan raised by candidate	Emphasis on relationship with Central government raised by candidate	Voting decision	Education level	Political Inclination
2016	Voting decision	Voting decision	Education level	Voting decision	Voting decision	Voting decision

Data Correlation

On election day, the dominant factors are

	Hong Kong Island	Kowloon West	Kowloon East	New Territories West	New Territories East	District Council (Second)
2008	Join 1st July Demonstration	Occupation	Why candidate attracts	Duration of being voter	Education level	/
2012	Channels of knowing candidates	Why candidate attracts	Voting decision	Reason of voting	Voting decision	Age

Manipulating the Data

- Three surveys:
 - 2008: 15,323 citizens
 - 2012: 16,253 citizens
 - 2016: 10,601 citizens
- Survey Question → Parameter
 - Each possible answer represented by a number (1, 2, 3..)
 - Same questions throughout years, different answer layout (Ugh!)
- Homogenize Data: (*Very boring and tedious*)
 - Result: 42,177 citizens total

Can now play with different models...

What is One Hot Encoding?

Cardinality:

- Education:
 - 0 - No Information
 - 1 - Elementary School
 - 2 - High School
 - 3 - Higher Education

One Hot Encoded:

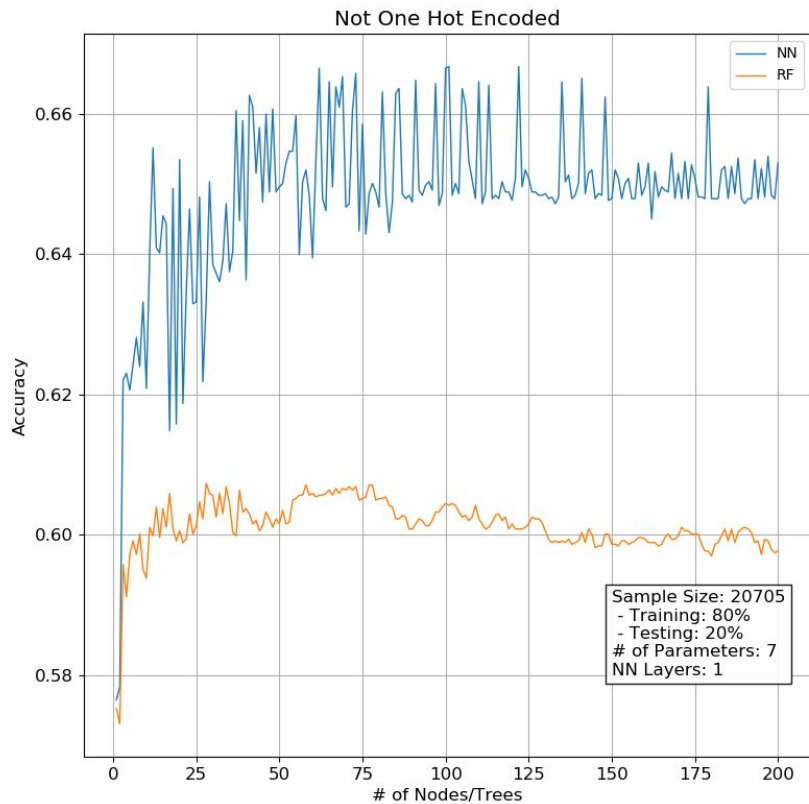
- Education
 - [0 0 0 0]
 - Index corresponds to answer!

Eg: [0 0 1 0]

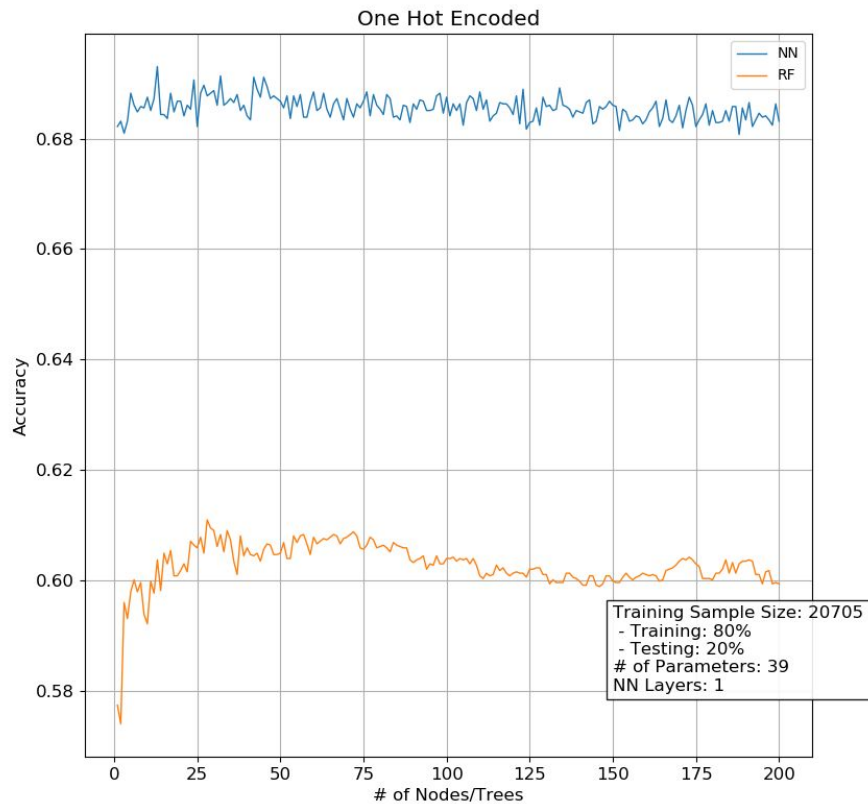
High School education

Evaluation of Algorithms

General Election, 2007 2012 Training



General Election, 2008 2012 Training



Legislative Council Election 2016

Total seats: 35

Total number of votes: 2,064,077

Pro-Choice: Pan-Democrats & Localists

Prediction Results 2016

Neural Network: 1 Layer, 15 Nodes

- Pro-Government: 48.14% of votes
- Pro-Choice: 51.56% of votes

Random Forest: 30 Trees

- Pro-Government: 37.80% of votes
- Pro-Choice: 49.62% of votes

Actual Results 2016

- Pro-Government: 871,016 (42.20%) votes
- Pro-Choice: 1,193,061 (57.80%) votes

Future Work

Improve prediction accuracy:

- Additional NN layers (DNN)
- Use of other machine learning algorithms and NN architectures
- Approach as a regression problem? (Predict probability of side winning)

Attempt to use model in other elections:

- Plenty of U.S. election data

Visualization Goals

Development of additional software to interface with Dark Matter as a plugin

Create 3D visualizations, and be able to interact with how the visualizations are displayed with graphical or scripted controls in Dark Matter

Data Visualization

-Still no access to DarkMatter

-Focusing on VR/Google headsets for prototyping of visualization



Figure 1: Projection of Hong Kong Map on to a sphere



Figure 2: View inside of sphere