Project Proposal

- Chenyang Jiang / cjiang77
- Haoran Teng / hteng22
- Chenhao Fang / cfang45
- Shikun Liu / sliu673
- Zixiang Xu / zxu475

Description

There is election data for the USA as a whole and there is also election data for every precinct within every state. (Americans vote for a specific candidate instead of directly selecting a particular political party. With a federal government, officials are elected at the federal (national), state and local levels. On a national level, the president, is elected indirectly by the people, through an Electoral College. In modern times, the electors virtually always vote with the popular vote of their state. All members of Congress, and the offices at the state and local levels are directly elected.) The data is totally more than 30GB.

URL：https://www.kaggle.com/paultimothymooney/open-elections-data-usa

Statistical question

Considering hillary and trump，whose followers are more inclined to vote early?

Or, statistically, is the difference of, whether followers are inclined to vote early, significant or not?

Short code snippet that reads the data onto our laptop

```
u=read.csv("D:/Desktop/archive/open-elections-data-by-state-and-precinct/open-elections-data-by-state-and-precinct/openelections-data-ak/2016/20161108__ak__general__10__precinct.csv")
```

(The data set is easy to read since all the data are stored seperately in csv files.)

Description of the variables available

District: the corresponding number of district in each states.

Office：the target position of the candidate, including 'President', 'State House', 'U.S. House', 'U.S. Senate' etc.

Party: which party is the candidate affiliated to.

Candidate: names of candidates.

Votes: how many votes a candidate received in one singal precinct.

Precinct：how the votes are received, including 'early voting' and 'absentee'.

Computational tools

Two proportion hypothetheis test in all states.

CHTC: we plan to use CHTC in order to process our large data set.

R: we can perform tests by running Rscripts.