**Log**

**Sat 8/10/19**

* **E: downloaded analysis file with following columns:**
  + **"state","county","fips"**
  + **"trump16","clinton16","otherpres16","romney12","obama12","otherpres12"**
  + **"demsen16","repsen16","othersen16","demhouse16","rephouse16","otherhouse16","demgov16","repgov16","othergov16","repgov14","demgov14","othergov14"**
  + **"total\_population"**
  + **"cvap" – citizen voting age population**
  + **"white\_pct","black\_pct","hispanic\_pct","nonwhite\_pct","foreignborn\_pct","female\_pct","age29andunder\_pct","age65andolder\_pc"**
  + **"median\_hh\_inc"**
  + **"clf\_unemploy\_pct" – as percent**
  + **"lesshs\_pct","lesscollege\_pct","lesshs\_whites\_pct","lesscollege\_whites\_pct"**
  + **"rural\_pct","ruralurban\_cc" (9 codes)**
* **T:**
  + **Created jupyter notebook and imported csv file (fixed some minor problems along the way)**
  + **Some basic analysis of data**
  + **Grouped by state…Mich, Pa, Wisc looked reasonable**
* **P: Brainstormed ideas for future development**

**Sun 8/11/19**

* **P: more brainstorming and planning**
* **E, T: downloaded file from Federal Elections Commission with electoral and popular vote by state and other data:**

**[https://transition.fec.gov/pubrec/fe2016/federalelections2016.xlsx](Federal Elections 2016 - Federal Election Commission - FEChttps://transition.fec.gov/pubrec/fe2016/federalelections2016.xlsx)**

* **E: copied tables from wikipedia re Maine 2016 voting**

**[https://en.wikipedia.org/wiki/2016\_United\_States\_presidential\_election\_in\_Maine]( https://en.wikipedia.org/wiki/2016_United_States_presidential_election_in_Maine)**

* **E: copied tables from a similar Nebraska wikipedia page**
* **E: reviewed census activities**
  + **unit 6**
    - **3-8 – by zip code: Zipcode,Population,Median Age,Household Income,Per Capita Income,Poverty Count,Poverty Rate (calculated)…probably available by county…also had a query for county name and state**
    - **3-9 – by state: State,Name,Population,Median Age,Household Income,Per Capita Income,Poverty Count,Poverty Rate (calculated), Unemploymernt Count , Unemployment Rate (calculated)…also a google map**
    - **3-10 – similar…more maps**
  + **unit 10**
    - **1-1 – data by city from sql database: 'CityState', 'city', 'state', 'Population', 'White Population', 'Black Population', 'Native American Population', 'Asian Population', 'Hispanic Population', 'Education None', 'Education High School', 'Education GED', 'Education Associates', 'Education Bachelors', 'Education Masters', 'Education Professional', 'Education Doctorate', 'Poverty', 'Employment Labor Force', 'Employment Unemployed', 'Employment Male Computer Engineering', 'Employment Female Computer Engineering', 'Median Age', 'Median Male Age', 'Median Female Age', 'Household Income', 'Income Per Capita', 'Median Gross Rent', 'Median Home Value', 'lat', 'lng']**
* **E: other MIT data to potentialy use**
  + **County Presidential Returns – 2000-2016**

**Mon 8/12/19**

* **E – saved electoral vote sheet to csv, removing all but the 51 rows for 50 states plus DC; added to Hawaii, Texas, and Washington state since 7 electoral vote went to others; need to check later that total is 538**
  + **Will add three columns in this csv to get total electoral votes by state**
* **E – MIT “County Presidential Returns” – 2000-2016 (previously dopwnloaded)**
  + **columns: year, state, state\_po (abbreviation), county, FIPS, office, candidate (or other), party, candidatevo, totalvotes (same for all candidates), version (date in yyyymmdd format)**
* **E – looked at obtaining other census data using Census Bureau API and / or python census package used in lessons for units 6 and 10 (see above); ultimately decided delay consideration of this until a later phase**
* **E – loaded electoral vote csv into python notebook for further processing**
* **E – loaded presidential history csv into python notebook for further processing**
* **P – brainstormed about design of user capabilities**

**Tues 8/13/19**

* **P – added brainstorming notes to this document**
* **T – translated electoral vote data**
* **T – started translation of pres\_hist data**
  + **Analyzed all fields that were NaN**
  + **deleted rows with “candidatevotes” equal to NaN or 0**
* **asked Q of Ysis: where to do aggregation: SQL, front end, back end; decision will consider fact that I will not be updating database on a continuous basis**

**Wed 8/14/19**

* **Finish translation of pres\_hist data**
  + **change “party” to “other” if NaN and candidate is “other”**
  + **fix “state\_po” and “FIPS” for 15 which were NaN**
  + **ignore “totslvotes” field (recalculate instead)**
  + **verified that only “totslvotes” field has NaN’s**
  + **group row for each year and county**
  + **calculate “totslvotes” field**

**Design of user capabilities**

* **data by county**
  + **presidential elections 2000-2016; with bins (will identify battleground states)**
  + **total population**
  + **%’s of population: voting citizens, ethnicity, age, education with bins**
  + **HH income: with bins**
  + **unemployment rate, with bins**
  + **urban / suburban / rural codes**
* **simulation** 
  + **select multiple rules for changes: shift %, turnout %**
  + **apply rules with greatest effect for these two by county**
  + **map color coding stats (ex, red, purple, blue)**
  + **list of changes in state electoral votes**
* **lots of plots: scatter, histogram / bar, pier, line**
* **voter turnout by county in heavily black counties, 2012 vs. 2016…create percentage field**
* **2012 vs. 2016 comparison**
* **$ college educated by county, 2012 vs. 2018**
* **additional %’s, ratios / calculations**
  + **trump / (trump + clinton), same for clinton**
  + **other %’s after analyze data (ex, % 30-64 year olds)**
  + **turnout: total votes / voting citizens**
  + **grouping of urban / suburban / rural codes after analysis**
  + **state / national totals / %’s**
  + **bins for all data items (ex, battleground states)**
  + **aggregate fields for 2000 to 2016 with total vote fields**
  + **total vote fields also for main database**
* **bar graphs with rep / dem / other with b ins on x-axis anf ability to filter by other bins and values**
* **bar graphs with rep / dem / other with 2000, 2004, 2008, 2012, 2016 on x-axis, also with filtering**

**To Do’s**

* **keep documenting steps above**
* **analysis translation**
* **load into SQL database**
* **review and refine: stuff below, user capabilities, notebooks**

**Planning ideas – first phase**

* **see ETL unit notes**
* **everything based on planned front end capabilities**
* **translate**
  + **check translation for all three jupyter notebooks**
  + **split up SQL tables into smaller blocks**
  + **Two notebooks: analysis and tranlate**
  + **check field definitions for analysis file, perhaps looking at sources:** [**https://github.com/MEDSL/2018-elections-unoffical/blob/master/election-context-2018.md**](https://github.com/MEDSL/2018-elections-unoffical/blob/master/election-context-2018.md)
  + **Create % fields and ratios**
  + **State level aggregations for % and other fields; also Trump states and whole nation**
  + **bins and %’s for various data fields at county and state level; first do basic bar graphs or histograms to see where there appear to be relationships**
  + **calculate trump / (trump + Clinton) %, same for Clinton**
  + **Put SQL data in normal form**
  + **check vote totals by state for two files; Alaska?**
* **check consistency of 2016 data in two places**
* **load**
  + **One big combine files or separate**
* **see guideline for project in slack message**

**Planning ideas – possible later phases**

* **Extract**
  + **more census data (see comment from 8/12/19 above)**
  + **other data on Excel spreadsheet**
  + **other MIT data – see list below**
  + **google other sources of data**
* **Later user capabilities**
  + **Other races**
  + **Statistical tests**
  + **Multiple regression**
  + **State maps with counties**
  + **User set bins**
  + **Users choose states**
  + **Drill down**
  + **Maine / Nebraska eception**

**MIT files**

**1. US Senate 1976-2018 – “1976-2018-senate.csv”**

**2. US President 1976-2016 – “1976-2018-president.csv”**

**3. US Primary Elections – 2018 – a repository - skipped**

**4. County Presidential Returns – 2000-2016 – “countypres\_2000-2016.csv”**

**- a dataset – 50528 rows**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **year** | **State** | **state\_po** | **County** | | **FIPS** | | **office** | |
| **2000** | **Alabama** | **AL** | **Autauga** | | **1001** | | **President** | |
| **candidate** | **Party** | **candidatevotes** | | **Totalvotes** | | **Version** | |
| **Al Gore** | **Democrat** | **4942** | | **17208** | | **20181011** | |

**5. State precinct -level returns 2018 – “precinct\_2018.csv”**

**5a. State precinct level returns 2016**

**6. US Senate precinct level returns 2016 - skipped**

**7. US President Precinct-Level Returns 2016 – “2016-precinct-president.cs**

**8. State office level returns 2016 – “stateoffices2016.csv”**

**9. Local precinct-level returns 2016 – “2016-precinct-local.=csv”**

**10. US House of Representatives Precinct-Level returns 2016 - skipped**

**11. state constituency level returns 2018**

**12. US House 1976-2018 - “1976-2018-house.csv”**

**13. US General Elections 2018 – Unofficial Returns – individual csv files for about 27 states,**

**14. US General Election 2018 – analysis dataset (not Alaska)**