**Project 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 “totalvotes” field (recalculate instead)**
  + **verified that only “totslvotes” field has NaN’s**
  + **group row for each year and county**
  + **calculate “totalvotes” field**

**Thurs 8/15/19**

* **created postgresql database and loaded “electoral” dataframe into it**

**Sat 8/17/19**

* **finished “pres\_hist” dataframe translation and load**
* **finished “analysis” dataframe translation and load**

**Tues 8/27/19**

* **analyzed some of fields in analysis table**
* **'state', 'county', 'fips'**
* **'total\_population', 'cvap'**

**Thurs 8/29/19**

* **cbsa data**
  + **downloaded data from OMB from** <https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-metro-and-micro-statistical-areas.html>
  + **shows counties in metropolitan statistical areas**
  + **loaded into jupyter notebook; had to open in VS Code, then save under a different name to solve problem initially encountered when trying to loads into jupyter notebook**
  + **created two data frames: (1) metro names for each CSBA code with population over 1,000,000 and (2) list of counties for each CSBA code**
* **added FIPS code to pres\_hist jupyter notebook**
* **analyzed more fields in analysis data**
* **ethnicity: 'white\_pct', 'black\_pct', 'hispanic\_pct', 'nonwhite\_pct', 'foreignborn\_pct' – need to add an “oth\_pct” category**
* **gender: 'female\_pct'**
* **age: 'age29andunder\_pct', 'age65andolder\_pct' – add new field**

**Fri 8/30/19**

* **'median\_hh\_inc', 'clf\_unemploy\_pct'**
* **education: 'lesshs\_pct', 'lesscollege\_pct', 'lesshs\_whites\_pct', 'lesscollege\_whites\_pct',**
* **'rural\_pct'**
* **'ruralurban\_cc' - // in analysis file since / is special character**
* **land area from census bureau**
  + <https://www.census.gov/library/publications/2011/compendia/usa-counties-2011.html#LND>
  + **file name “LND01 2”**
  + **saved as csv file from Excel**
  + **imported into Jupyter notebook**
  + **used field “LND110210 base on checking three Alabama counties since this figure does not include water**
  + **excluded national and state records for which STCOU code ends in “000”**
* **design of Phase 1 user capabilities and project planning**

**Early September**

* **continued design and planning**
* **found a definition of US regions**

<https://www.google.com/search?q=us+regions+map&tbm=isch&source=iu&ictx=1&fir=WcEYLJcLDJXbcM%253A%252C-OcHhWEoedXmnM%252C_&vet=1&usg=AI4_-kTes1_dPCXv0r3oh44bRoJg0TNEdw&sa=X&ved=2ahUKEwjvoeG296_kAhVQIjQIHRPeCtAQ9QEwBHoECAUQDA#imgrc=WcEYLJcLDJXbcM:>**)**

**Thurs 9/5/19**

* **finished design and plan for Phase 1**

**Fri 9/6**

* **dropped all tables from SQL database in pgadmin4**
* **electoral votes by state: reviewed, reran, checked there when refresh pgadmin4**
* **historical presidential votes by county**
  + **added state abbreviation**
  + **same steps**
* **analysis data by county – same steps**
* **land area by county – same steps**
* **states**
  + **manually created state file with abbreviation, name, and region**
  + **used early September link in project log for regions**
  + **same steps**
* **metro areas**
  + **finished creating it**
  + **same steps**
* **jupyter notebook to create a single table by county**
  + **documented with markdown explanations in combined.ipynb file**
  + **same step**

**Sun 9/7/19**

* **starting from most recent homework, programmed index.html, app.js, and app.py to show grouped bar charts with 2016 voting totals**

**Mon 9/8/19**

* **planning**

**Tues 9/9/19**