**To do**

* **save files as app909.py, etc.**
* **separate function for creating bins**
* **item specific bins – median\_hh\_inc, pop / land area (just two for now)**
* **add other column**
* **totals rep/dem/oth for each chart**
* **no Jumbotron**
* **ability to select states**
* **ability to select other data items**
* **proof code: comments, variable names**
* **check vote totals: any counties missing? Alaska and Hawaii? Vs. electoral spreadsheet?**
* **how to handle following calculations? Add to combined database?**
  + **voting age citizens as % population**
  + **2000-2016 total votes as % cvap (i.e., voter turnout)**
  + **population / land area**
  + **rep / (rep + dem) %, same for dem, also as % total votes**
  + **nonwhite: < high school, < college, college**

**Phase 1 design**

* **list of tables and fields**
  + **states data: abbreviation, name, region name**
  + **metro areas**
    - **metro code, metro name**
    - **county FIPS code, metro CBSA code**
  + **analysis data by county**
    - **county data: name, FIPS code, state name**
    - **population: total, voting age citizens**
    - **ethnicity: white, black, hispanic, other, nonwhite, foreign**
    - **gender: female**
    - **age: <= 29, 30-64, >=65**
    - **economic: household income, unemployment**
    - **education: < high school, < college, college for all and white (so also have nonwhite)**
    - **urban / rural: rural %, ruralurban code**
  + **historical presidential votes by county**
    - **dem / rep / other / total for 2000, 2004, 2008, 2012, and 2016**
  + **land area by county**
    - **county FIPS code, land area**
  + **electoral votes by state**
    - **state abbreviation, electoral votes**
  + **calculated fields**
    - **voting age citizens as % population**
    - **2000-2016 total votes as % cvap (i.e., voter turnout)**
    - **population / land area**
    - **rep / (rep + dem) %, same for dem, also as % total votes**
    - **nonwhite: < high school, < college, college**
* **phase 1 front end**
  + **bar charts**
    - **for state, region, nation**
    - **bins on horizontal axis**
    - **bars for rep / dem / oth vote totals**

**Later to do’s**

* **formatting**
* **metro – Boston, other for a state**
* **calculate state, region, national and metro aggregations? Jupyter notebook?**
* **check “combined” jupyter notebook to make sure did not lose any data during merges**
* **for %’s, use hierarchy concept, such as dem as % both rep/dem and tot?**
* **2D tables with different bins on x- and y- axes and vote totals or other aggregations in cells; side by side with a chart re one of columns or rows**
* **“what if” simulations**
  + **user specified turnout and / or vote shifts rules by “bin” category**
  + **use highest for a county when more than one turnout rule applies; same for vote shift rules**
* **national map with red / blue color coding by state**
  + **showing who won 2016, other years, or a “what if” simulation**
  + **also show electoral votes totals**
  + **state-specific info when mouse over**
* **creative way to compare different years and / or “what if” simulations**
  + **bar graphs**
  + **line graphs**
  + **side-by-side national maps highlighting differences [flashing?]**
* **able to filter everything: by state, by bins, other values, convenient “drill down” capabilities**
* **other visualizations: scatter, histogram / bar, pie, line [and other stuff we have learning in Matplotlib, U15 on]**
* **other plots: scatter, histogram / bar, pie, line**
* **check vote totals by state for two files, ex, Alaska**
* **grouping of urban / suburban / rural codes**
* **data over time (ex, analysis data)**
* **check consistency of 2016 data**
* **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)
* **see ETL unit notes**
* **notes from TV lady**
  + **Iowa, NC. Arizona, Florida will be tossups**
  + **Rachel Bitcofer**
  + **key is turnout Dem, Ind, not switching**
  + **“It’s all about the base”**
* **Extract**
  + **more census data (see comment from 8/12/19 above)**
  + **other data on Excel spreadsheets**
  + **other MIT data – see list below**
  + **google other sources of data**
* **Later user capabilities**
  + **Other races such as senate, house, etc.**
  + **Statistical tests**
  + **Multiple regression**
  + **State maps with counties**
  + **User set bins**
  + **Users choose states**
  + **Maine / Nebraska exception**

**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)**