**Katie’s dataset notes**

* 01.Overview\_Geospatial\_Data.pdf: Changes recommended.
  + Prose notes because I don’t have the original slide files.
  + Title: I’d make the period in the title into a “\_”.
  + Slides 17-18
    - Location example looks like an all-white ranch/4H program (<https://asotincountyfairandrodeo.org/4-h/>)
    - Attribute explanation not clear (add example attributes?)
    - Possibly use this location instead? <https://commons.wikimedia.org/wiki/File:Coyote_Point_Trail_at_Whitewater_State_Park,_Minnesota_(44136078811).jpg> (photo has a CC 2.0 license)
  + Slide 31: CA maps look great (local, social science-y)
  + Slide 34: Given the racialization of what counts as “crime” in the United States (and even more overt racism in police presence), I’d replace crime locations with something else, e.g., motor vehicle accident locations or point environmental pollutant emissions locations from the Toxics Release Inventory (e.g., <https://www.epa.gov/toxics-release-inventory-tri-program/tri-basic-data-files-calendar-years-1987-2019>)
  + Rest of slides: Good data!
* 02\_Introduction\_to\_GeoPandas.ipynb: No changes recommended
  + Datasets used (all from United States Census):
    - California county shapefiles: 'notebook\_data/california\_counties/CaliforniaCounties.shp'
      * Social science: Census county boundary data likely used by a lot of quantitative social science researchers
      * Local: Fits CA/Berkeley D-Lab theme
    - California places shapefiles: 'notebook\_data/census/Places/cb\_2018\_06\_place\_500k.zip'
      * Social science: Census county boundary data likely used by a lot of quantitative social science researchers
      * Local: Fits CA/Berkeley D-Lab theme
* 03\_Introduction\_to\_GeoPandas.ipynb: No changes recommended
  + Datasets used
    - ‘notebook\_data/california\_counties/CaliforniaCounties.shp’
    - ‘notebook\_data/us\_states/us\_states.shp’
    - ‘notebook\_data/census/Places/cb\_2018\_06\_place\_500k.zip’
  + Same comments as for 02\_Introduction\_to\_GeoPandas.ipynb
* 04\_More\_Data\_More\_Maps.ipynb: No changes recommended to datasets, some to language around housing parcels
  + Datasets used
    - 'notebook\_data/transportation/BerkeleyBikeBlvds.geojson' (Berkeley bike boulevards)
      * Social science-relevant
      * Local
    - 'notebook\_data/alco\_schools.csv' (Alameda school locations)
      * Social science-relevant
      * Local
    - 'notebook\_data/parcels/parcel\_pts\_rand30pct.geojson'
      * Social science: Housing parcel data is definitely social science-relevant!
      * Local
      * Wasn’t clear on first read what a parcel was or the origin of the (I’d say “housing” or “real estate” parcel data and clarify the origin of the dataset in the workshop text)
    - ‘notebook\_data/berkeley/BerkeleyCityLimits.shp’
      * Social science-relevant
      * Local
* 05\_Data-Driven\_Mapping.ipynb: No changes recommended to datasets (same as in 04)
  + Datasets used
    - 'notebook\_data/california\_counties/CaliforniaCounties.shp'
    - 'notebook\_data/alco\_schools.csv'
    - 'notebook\_data/transportation/BerkeleyBikeBlvds.geojson'
* 06\_Spatial\_Queries.ipynb
  + Datasets used
    - 'notebook\_data/census/Tracts/cb\_2013\_06\_tract\_500k.zip'
    - 'notebook\_data/protected\_areas/CPAD\_2020a\_Units.shp'
    - 'notebook\_data/berkeley/BerkeleyCityLimits.shp'
    - 'notebook\_data/alco\_schools.csv'
    - 'notebook\_data/transportation/BerkeleyBikeBlvds.geojson'
    - 'notebook\_data/transportation/bart.csv'
* 07\_Joins\_and\_Aggregation.ipynb
  + Datasets used
    - 'notebook\_data/census/ACS5yr/census\_variables\_CA.csv'
    - 'notebook\_data/census/Tracts/cb\_2013\_06\_tract\_500k.zip'
    - 'notebook\_data/alco\_schools.csv’
* 08\_Pulling\_It\_All\_Together.ipynb
  + Datasets used
    - 'outdata/tracts\_acs\_gdf\_ac.json'
    - 'notebook\_data/other/ca\_grocery\_stores\_2019\_wgs84.csv'
* 09\_ON\_YOUR\_OWN\_A\_Full\_Workflow.ipynb
  + Datasets used, no changes recommended, but not listed at beginning
    - 'notebook\_data/ac\_voting\_locations.csv' (polling station locations)
    - ‘notebook\_data/census/Tracts/cb\_2013\_06\_tract\_500k.zip’ (census tract boundaries)
    - ‘outdata/berkeley\_places.shp’ (Berkeley places also from census)
    - Bus routes: 'notebook\_data/transportation/Fall20Routeshape.zip'
* 10\_OPTIONAL\_Fetching\_Data.ipynb
  + Datasets used:
    - Census Data: <https://www2.census.gov/geo/tiger/GENZ2018/shp/cb_2018_06_tract_500k.zip>
    - SF Bikeway Data: <https://data.sfgov.org/api/geospatial/ygmz-vaxd?method=export&format=GeoJSON>
    - Berkeley Bikeway Data: <https://data.cityofberkeley.info/api/geospatial/fgw9-98ic?method=export&format=GeoJSON>
    - OSMNX Library SF and Berkeley Data
* **11\_OPTIONAL\_Basemap\_with\_Contextily.ipynb**
  + **Datasets used (**not listed at beginning)
    - OpenStreetmap Mapnik basemap (via contextily)
    - 2019 cartographic boundary (cb\_) file of California (06) places: <https://www2.census.gov/geo/tiger/GENZ2019/shp/cb_2019_06_place_500k.zip>
    - BART station locations: 'https://raw.githubusercontent.com/dlab-berkeley/Geospatial-Fundamentals-in-Python/master/notebook\_data/transportation/bart.csv'
    - ax, source=cx.providers.NASAGIBS.ModisTerraTrueColorCR (via contextily)
* **12\_OPTIONAL\_Interactive\_Mapping\_with\_Folium.ipynb**
  + **Datasets used (**not listed at beginning)
    - notebook\_data/bartmap\_example.html
    - OpenStreetMap tileset
    - CartoDB Positron tileset
    - Stamen Watercolor tileset
    - "notebook\_data/california\_counties/CaliforniaCounties.shp"
* **13\_OPTIONAL\_geocoding.ipynb**
  + **Datasets used (**not listed at beginning)
    - **Alameda County schools: "./notebook\_data/alco\_schools.csv"**
    - **Contextily basemap ‘ax’**
    - **CartoDB Positron tileset**
* **14\_OPTIONAL\_Plotting\_and\_Mapping\_with\_Altair.ipynb**
  + **Datasets used (**not listed at beginning)
    - ACS 5-year data: notebook\_data/census/ACS5yr/ (especially Alameda County)
    - Census tract shapefiles: 'zip://./notebook\_data/census/Tracts/cb\_2018\_06\_tract\_500k.zip'
* **15\_OPTIONAL\_Voronoi\_Tessellation.ipynb**
  + **Datasets used (**not listed at beginning)
    - 2020 General Election voting locations for Alameda County: 'notebook\_data/ac\_voting\_locations.csv'
    - Census tracts for Alameda County: "zip://notebook\_data/census/Tracts/cb\_2013\_06\_tract\_500k.zip"
* **16\_OPTIONAL\_Introduction\_to\_Raster\_Data.ipynb**
  + **Datasets used (**not listed at beginning)
    - Missing dataset list at the top
* **99\_Questions\_Answers.md: no datasets!**