* Top tasks-
  + Use CDC SVI instead of HPI?
  + Check CHSI geocoding
  + CHECK 2013 or 2015 ACS data as denominator for 2011-2015
  + Check/decide DOF data
  + Narrower sidebar panel?
  + Bug tracker
  + Two maps
  + Two ranks
* Key definitions
  + “Communities”—throughout the CCB, communities are defined by “Medical Service Study Areas”, a unique California geographic designation, based on aggregation of census tracts, constructed by the California Office of Statewide Planning and Development (OSHPD) with each decennial census. REF. MSSAs provide the CCB with a good surrogate for “communities” because: (1) there are 5XX MSSAs for the 2010 census, providing much more geographic granularity than the 58 California counties, and much greater numerical/statistical stability than the 8000+ California 2010 census tracts, and (2), as much as possible, they are aligned with “communities” in the important sense of geographic, cultural, and sociodemographic similarities—this is generally more true for urban then rural MSSAs, because of the larger size of MSSAs in rural areas, and (3) the names associated with each MSSA has some resonance in many cases with “community” as noted above. Although not yet implement in the CBD in a fully automated fashion, users can work with the CCB project team to provider their own list of communities—based on designated census tracts---and these can easily be brought into a local, or even statewide implementation of the CCB.
  + Social Determinants of Health—
* Data and other key inputs:
  + Death data
    - Provided by California Department of Health (CDPH), Center for Health Statistics and Informatics (CHSI)
      * Xxxdeath files, 2001 to 2015 URL
      * Data coding and cleaning issues
      * A death record was considered to be of a California resident based on XXX. A tiny fraction of these records geocoded to locations outside of California, and others had anomalies suggesting the possibility that the residence was not California. Efforts are ongoing to address these issues. However, the number of such anomalies are so relatively minuscule, that they are extraordinarily unlikely to have any impact on observed patterns and trends.
      * Prior to CCB project receipt of these death data, CHSI staff geocoding the data using XYZ. For the years where the CCB uses these data for census tract and (and therefore community) geographies, 2011-2015, a high percent of records geocoded to a valid census tract (9X.X% to 9X.X%)—the remaining records contained invalid addresses and/or other anomalies. CHECK WITH CHIS—CENSUS TRACTS FOR PO BOXES WHERE NOT USED? WHAT GEOCODE SCORE WAS USED TO INCLUDE RESULT?
      * County based on XXX=
  + Social Determinants of Health (SDOH)
    - The CCB currently contains a small, exploratory of SDOH variables extracted from the publix XXX files from the Healthy Place Index URL;XXX CHECK CDC SOCIAL VURNURABILITY INDEX short term road-map plan to extract directly from US Census / American Community Survey API (URL) using the R tidycensus package (URL)
  + Population data
    - For census tracts (and therefore communities) population denominator data are based on the American Community Survey 5-year extract centered around the 5-year tract/community data being analyzed in the CBD (i.e. current 2011-2015 death data use the 2013 ACS data, being the center point of 2011-2015.) ETC. Communities (MSSA).
    - ACS data are extracted directly from the Census/ACS API (Application Program Interface) using the R tidycensus package REFURL.
    - For counties, population denominator data are based on California Department of Finaces (DOF) estimates….
* GIS
  + Boundry (or “shape”) files for CBD were generated using the R XXX package, modofiy to be smaller using the xx function of xx, and with removal of physical islands of the west coast of some counties using a custom xxxxx.
  + A shape files and GIS objects are stoared and manipiuated as “simple features” objects XXXX
  + All maps are projected based on XXX projection (although this can be easily changed as needed).
* ICD-10 Mapping
  + In the current version of the CBD project, only the single underlying cause of death ICD-10 code is use to classify the cause of death. Efforts are underway to incorporate ‘multiple cause of death” codes for some conditions in a future release of the CBD, and based on user interest/support.
  + We used a variate of the WHO/GBD condition list hierarchical outline to create our own hierarchical outline with three levels. The first level includes “Infectious Diseases”, “Corany Heart Disease”, “Cancer/Malignant Neoplasms”, “Other Chronic Conditions”, and “Injury” as well as all cuases combined. For data displayed at the census tract level, only this level of the hierarch is included. The next level, our “Public Helath Level” breaks each of each of the top level down into a total grouping of XXX condtions; the final detailed level breaks a few of these PH level conditions down further, for a total of XXX condtions, which is dispaleyed at the county lelvel.
  + The primary basis for the ICD-10 to condition mapping is the WHO/GBD table from XXXX found here URL ref. For the primary mapping, we were not able to use the IHME/GBD mapping system as descreibed in their recent publications (e.g. HEREURL), because that system results in approximately XX% of (California) death being mapped to “garbage codes”, for which more sophisticated, and possibly not apprpaorte for our system, methods would need to be employed. However, to enhance or use of the WHO/GBD system we did compared the mapping of XX,XXX,XXX California deaths based the two systems. As a results of these comparisons we did find a number of instnaces where the two systems devidated meaningfaully in their mappings, and for which we believed we could improve the WHO/GBD “mapping” by using the IHME/GBD mapping. All of these modifcations are carefully described in a key resoruces tool for the CBD, avaialbe hereREF.
  + In addition, becase of our focus on the “Public health” list of condtions, we remapped a number of ICD-10 codes from the GBD mapping to our own CBD system. These include alcohol… All of these modifcations are aslo carefully described hereREF
* Census Tract Data Data Isses
  + XXX census tracts included in some designations of California 2010 census tracts were excluded for all project calculations and maps for communities or census tracts because they are listed as containing 0 population.
  + Another X tracts with population less than X were also excluded because of the limited utility of any data from these tracts.
  + Census tracts (and communities) where greater than X percent of the population live in congreagant living quarters ARE/WILL BE noted with an \* on relevant maps and charts. For some comparisons (e.g. of rates) these tracts could be removed from the larger geographies in which they are contained, based on user request.
  + Based on detailed review of multiple data soruces, we observed a number of instances where stated county of resisnce was not consistent with the census tract to which that death geocode—in these instances we “recoded” the county based on the addressed and subsequent geocode.
* Formulas and measures
  + Years of Life Lost (YLL)
    - Following the methods of the Global Burden of Disease Study, the YLL for each death is based on the age at death, and the additional number of years a person living in an optimal setting could be expected to live. REF. For example someone dying at birth would be associated with 9x.x YLL, someone dying at 25 associated with 7.x years, and someone dying at 98 would with 2.x years. Beyond the published data, we associated 1.0 YLL for anyone dying above age X.
    - Our mapping of age at death to YLL can be found HEREURL.
  + Crude rates
    - All rates are expressed as per 100,000 population based on 100,000\*(number of events (e.g. deaths) / midyear population
    - Confidence intervals for crude rates are based on the XXX methods as implemented in the epitools package of R REFURL.
  + Age adjusted rates
    - Age-adjusted rates are based on the “direct” method, using the standards definitions and procedures. Great descriptions, and the motivations for these methods, found here URLREF.
    - XXX Age-grouping were used for these calculations
    - The 2000 Standard Population from NCHS/CDC was used
    - The age group mapping at this standard population data can be found here URL.
    - The age-adjusted calculation, and generation of confidence intervals was conducted using the ageAdjust.Direct() function of the R epitools package. URLREF.
    - Because a very small number of census tracts with otherwise useful data, had 0 population in one or more age strata (often the youngest or oldest strata, often for just one sex), the above mentioned function was modified such that rates in such strata were assigned to (reasonably enough) be 0 (rather than undefined/infinity), allowing an adjusted rate to be calculated.
  + Life expectancy
* Other Key References
  + Global burden of disease
  + R tools and packages