# DO NOT ADD any punctuation except commas and periods, hash mark for R comments, and single and double quote using the same exact character set as below

# textIntroA # measureHelp

# textIntroB # causeHelp

# textIntroC # statecutHelp

# textNote.real # cutmethodHelp

# textNote.fake # SDOHHelp

# mapTab

# conditionTab

# conditionTableTab

# condtionSexTab

# rankGeoTab

# trendTab

# sdohTab

#--Home Page Side Bar Text ------------------------------------------

textIntroA <- "The California Community Burden of Disease Engine (CCB) is a tool to explore data on burden of disease in multiple levels of geographic granularity."

textIntroC <- " The CCB currently displays over 15 years of California condition-specific mortality burden data, using a range of measures displayed at the statewide, county, community, and census tract levels, with interactive rankings, charts, maps and trend visualizations."

textNote.real <- "This version of the app contains features that are still being tested and refined. While the contents have been reviewed extensively, the tool likely still contains 'bugs' and may generate error messages based on some selection combinations. While our team is continuously working to address any identified issues, we are confident the application in its current form will, overall, provide important and reliable insights. We look forward to and appreciate any observations you make regarding bugs, errors, or anything else of concern. We also welcome supportive comments."

textNote.fake <- "NOTE: THIS VERSION OF THE ENGINE IS FOR DEMONSTRATION PURPOSES ONLY - THE DATA ARE NOT REAL - THEY ARE A RANDOM SUBSET OF RANDOMLY DISTORTED DATA"

#--Home Page Main Panel ------------------------------------------

above1 <- paste0("<ui style=align='left'><b>

California Community Burden of Disease and Cost Engine (CCB):

</b>

<br>

<h4><b><i>

An emerging toolset for epidemiologic analysis and scientific insight, exploring the intersection between health disparities and place

</b></i></h4></ui>")

#--Warning Messages Etc. -----------------------------------------

tractWarning <- "NOTE: AT THE CENSUS TRACT LEVEL, ONLY THE 'TOP LEVEL' CONDITIONS CAN/WILL BE DISPLAYED; ALSO, RENDERING STATE-WIDE TRACT-LEVEL MAPS CAN TAKE MANY SECONDS--PLEASE WAIT"

oshpdModal <- "<b> The presentation of these hospitalization data should be considered preliminary. There are a number of nuances and considerations relevant to interpreting these charts noted in the technical documentation. Of particular note, the summaries of 'charges' shown, are specifically, <u>charges</u> based on hospital administrative systems, and do not indicate the actual costs, reimbursements, or payments for those charges. However these charge data provide one lens of valuable information regarding the patterns of monetary burden. We are also in the midst of assessing optimal ways of grouping conditions/diseases and other aspects of the sharing of these data. We welcome your input. </b>"

# DROP DOWN HELPS --------------------------------------------------

levelHelp <- paste0('<li style="margin-left: 40px">The <b>Top Level</b> includes only the five most aggregated conditions (e.g. Cardiovascular diseases, Injuries) and All Causes combined.</li>

<li style="margin-left: 40px">The <b>Public Health Level</b> includes a list of about 60 conditions. This Level provides greater clinical detail and public health program specificity than the Top Level. </li>

<li style="margin-left: 40px">The <b>Detail Level</b> includes a small number further detailed conditions (e.g. Non-Hodgkin lymphoma, Opiod Use Disorders, Falls) <b>AND</b> includes all conditions in the Public Health Level. </li>

')

dxGroupsHelp <- paste0(

'<li style="margin-left: 40px">

<b>Global Burden of Disease (GBD)</b> groups the ICD-10-CM codes into conditions based on our modified version of the Global Burden of Disease system (describe in detail in the technical document), but currently includes ONLY "high volume" conditions and a sample of other conditions of clear programmatic public health interest in California. For these hospitalization data this system is <b>not currently exhaustive</b>.

</li>

<li style="margin-left: 40px">

<b>Major Diagnostic Categories (MDC)</b> group principal diagnoses into 25 mutually exclusive diagnosis categories. The categories correspond to a single organ system or etiology and, in general, are associated with a particular medical specialty.

</li>

<li style="margin-left: 40px">

<b>Medicare Severity Diagnosis Related Group (DRG)</b> categorizes patients according to clinical coherence and expected resource intensity. The assignment of a DRG is based on: principal diagnosis and any secondary diagnoses, procedures performed, comorbidities and complications, age of patient and sex, and discharge status.

</li>

')

measureHelp <- paste0('<p>The current MEASURES of deaths are:</p>

<ui style="list-style-type:circle">

<li style="margin-left: 40px">Years of Life Lost (YLL)</li>

<li style="margin-left: 40px">Years of Life Lost Rate (per 100,000 population)</li>

<li style="margin-left: 40px">Age-Adjusted YLL Rate</li>

<li style="margin-left: 40px">Number of Deaths</li>

<li style="margin-left: 40px">Crude Death Rate (per 100,000 population)</li>

<li style="margin-left: 40px">Age-Adjusted Death Rate</li>

<li style="margin-left: 40px">Mean Age at Death</li>

<li style="margin-left: 40px">Standard Mortality Ratio (SMR)</li>

</ui><br>

<p>No one measure is best--each measure provides a different view or perspective into the impact of the condition. For example, <b>Number of deaths</b> is the simplest, most direct measure, and has clear intuitive meaning, and, other things being equal, will be larger in areas with larger populations. <b>Crude Death Rate</b> takes the size of the population into account by dividing the number of deaths by the number of people in the population (multiplied by 100,000 for interpretability, i.e. number of deaths per 100,000 people).

<b>Age-adjusted Death Rate</b> is the rate that would have existed if the population had the same age distribution as a reference population. This allows for comparisons between populations with differences in age distributions, accounting for the fact that age itself is generally correlated with higher mortality.

<br><br>

<b>Years of Life Lost</b> sums all the years of life prematurely lost across all people that die from that condition, and is influenced by the age at which people die from the condition and the number of people that die from that condition. <b>Years of Life Lost Rate</b> divides the YLL sum by the number of people in the population, and then multiplies by 100,000 for interpretability. It is the YLL equivalent of the Death Rate. <b>Age-Adjusted YLL rate</b> adjusts for the age structure of the population, and is the YLL equivalent of the Age-Adjusted Death Rate.

<br><br>

<b> Average Age at Death</b> directly measures the mean age of death for a particular condition and/or for a particular geography. Younger average age at death has intrinsic greater impact on a population, and contributes to larger YLL.

<b> Standard Mortality Ratio (SMR)</b> shows the county rate of a condition divided by the rate for the condition in the State overall. This measure will highlight counties that have especially high (or low) rates of a condition compared to the State rate, even if the condition does not have a large number of deaths.

')

causeHelp <-

'The <b>Cause</b> list contains about 70 conditions, organized based on the hierarchal structure of the Global Burden of Disease Project of the <a href=" https://www.who.int/healthinfo/global\_burden\_disease/about/en/""> World Health Organization (WHO)</a> and the <a href=" http://www.healthdata.org/gbd/about "">Institute for Health Metrics and Evaluation (IHME)</a>. We have made modifications to the WHO and IHME lists to enhance the usefulness and applicability for U.S., and specifically California, public health priorities and programs. These modifications are described on the Technical Tab of this website.

<br><br>

Depending on the geographic level selected, different levels of detail in the cause list will be available. The more granular the geographic level, the less granular the cause list will become, for statistical stability. [For our partners in Local Health Departments and in other California Department of Health Programs, fully granular data on all levels are available on request.]

<br><br>

<b>Top Level</b> conditions are a mutually exclusive and exhaustive set of five high-level conditions indicated by capital letters: A) Communicable, maternal, perinatal, and nutritional conditions; B) Cancer/Malignant neoplasms; C) Cardiovascular diseases; D) Other chronic; E) Injuries; and Z) Unknown/Missing value

<br><br>

<b>Public Health Levels</b> are a mutually exclusive and exhaustive list of about 60 conditions that are subcategories of the Top Level conditions; they are indicated by numbers associated with the capital letter of the Top Level condition under which they are subcategories. This level provides greater clinical detail and public health program specificity.

<br><br>

<b>Detail Levels</b> are available for a small number of the Public Health Level conditions; these are indicated by a lowercase letter and allow for further specificity.

<br><br>

Our full cause list hierarchy is available <a href="gbd.ICD.MapIMAGE.pdf"">here</a>. Additional information, including a link to the list of detailed ICD-10 codes as they map to all conditions, can be found on the Technical Tab of this website.

<br><br>

We welcome comments and suggested changes to any aspect of our hierarchal list.

'

stateCutHelp <-

"The <b>State-based cutpoints</b> button changes the way the Measure is broken down or grouped for display in the map and the legend. If the box is checked, the cut-points are based on the data for all 58 counties for the most recent five years. If the box is unchecked, the cut-points are based on just the county being displayed in the current map.

<br><br>

Looking at the map with the box checked allows you to see how things compare to the State overall. Looking at the data with the box unchecked allows you to see how things compare within the selected county. For example, if most communities in a given county have high rates of something compared to other counties, that county would be almost all red (indicating high rates) with the box checked, and would show a more varied distribution with the box unchecked.

<br><br>

Different insights can be gained by looking at these maps in the two different ways.

"

cutmethodHelp <-

"Specifies method used to determine the cut-points for the color categories. <b>Quantile</b> divides the frequency distribution into equal categories, each containing the same fraction of the total dataset. <b>Fisher</b> uses the Fisher-Jenks algorithm which reduces the variance within categories and maximizes the variance between categories."

# ---- TAB Help ---------------------------------------------------

mapTab <- paste0("

These maps display the geographic distribution of disease burden among counties and communities across California. The <b>Geo Level</b> options allow the user to change the display from county, to community, to census tract. This selection is one of the key concepts behind the app, that place matters! Insights into the burden of disease must be explored at multiple geographic levels, especially granular community levels. Data at the community and the census tract levels are aggregated to 5-year intervals.

<br><br>

Users can select either the state as a whole or zoom to a specific county for a better view of just that county, and its subcounty detail.

<br><br>

Users can select from various measures of mortality to assess burden of disease.

Selecting the <b>State-based cutpoints</b> option allows for comparisons based on the statewide distribution instead of just within the county.

<br><br>

The interactive map allows for zooming in and out to see streets or other geographically identifying locations.

Also, the interactive map has a pop-up which display information for the geography selected.

<br><br>

The static map is generally better for using in an external presentation (e.g. pasting into a PowerPoint), since it is 'crisper' without the background map layer. The Place Names option displays county and community names.

")

conditionTab <-

"This tab displays cause-of-death rankings for either a selected county or the whole state. The figure shows the ranking based on five different measures, and can be sorted based on any of these measures using the <b>Measure Sort Order</b> tool. Different insights can be gained by ranking on different measures (e.g. ranking on 'YLL per 100,00 population' highlights overall burden; ranking on 'mean age at death' shows the conditions that impact young people the most, and ranking on 'Standard Mortality Ratio (SMR)' show those conditions for which a county has particularly high rates compared to the State average). The <b>Levels to Show</b> option allows users to select between broad or narrow categories of conditions. The <b>How Many</b> button determines how many causes of death to display on the graph.

<br><br>

Ranking on SMR provides a special window in the potentially unique priority of a condition in the selected geography. A large SMR means the condition is especially high in that geography relative to the State average, even if the condition does not have a large number of deaths. This is an important way to detect conditions that, while perhaps not common, are unusually high (or low) in a county or community in which one is interested. To aid in using this measure, the vertical red line is at 1.2, corresponding to 120% of the State average. The green line is at 0.8, 80% of the State average and the grey line is at 1.0, right on the State average.

<br><br>

Because the 'base' ratio for the SMR is the state rate, the SMR is not shown when the statewide 'California' geography is selected; it would be 1.0 in all cases.

<br><br>

"

conditionTableTab <-

"This is a tabular version of the Rank Conditions tab, providing for a more granular examination of specific numbers or rates. Users can sort the table on any of the measures and can use the search window allows users to quickly find a specific condition.

"

conditionSexTab <-

"(This tab work in progress, awaiting further development of the tab)

<br><br>

This tab ranks causes within a selected geography separately for males and females. It can highlight conditions that appear to be a leading cause of death for one sex but not the other.

"

rankGeoTab <-

"This tab displays the ranked order of counties in California or the communities within a selected county for a selected condition. These rankings highlight places where a particular condition is the highest as well as highlighting geographical disparities of the condition. Years of life lost and number of deaths will tend to be highest in areas with the largest populations, whereas rate measures adjust for population size.

<br><br>

Note that higher ranking counties or communities may not be meaningfully higher from a statistical perspective; examining the confidence intervals will help determine if there is a meaningful difference or not. Users can display confidence intervals by checking the 95% CIs option. In the current version, confidence intervals are only displayed for the crude death rate and age-adjusted death rate but will be available soon for other measures.

"

trendTab <-

"This graph displays the trend over time for a particular condition within a selected geography, separately for males, females, and the total population. Reviewing the trend over time is important for understanding which problems are improving and which are getting worse.

<br>

Note: Because the data for the communities and census tract are currently aggregated for 5 years, those data are not available currently in the trend tab.

"

sdohTab <-

"This tab is preliminary and under development.

<br><br>

This scatter plot displays the correlation of a selected <b>Social Determinant of Health</b> measure with a selected condition. Each dot maps the value of the social determinant measure against the value of the condition measure for one geographic unit (county, community, or census tract). Because this association is 'ecologic' (correlation of geographic units, not of individuals), it is particularly important in this tab to look at measures that take into account the size and age distribution of the population, such as age-adjusted YLL rate and age-adjusted death rate. While correlations do not indicate causation, they are a potentially important way to understand the differential roles of some social determinants of health on disease outcomes.

<br><br>

In the current version, the colors represent the regions of the state; soon the colors could represent the rurality levels of the county/community/tract or other factors.

<br><br>

In the current version, the size of the dots is proportional to the size of the population represented by the dot. This too, could represent other things.

<br><br>

Note: Currently this tab only displays one variable, but the display and analysis in this tab will be expanded to include multiple variables simultaneously.

"

lifeExpectancyTab <-

" This chart shows the estimated Life Expectancy for California overall and for most California counties, for males and females, from 2005 to 2018, and includes confidence intervals. Life Expectancy is a familiar and widely used measure, which summarizes in one number the 'force of mortality' in a population, and provides a valuable single measure to compare the overall health status between populations.

<br><br>

'Life Expectancy', or technically 'Life Expectancy at Birth', is calculated based on the number of people in each age group that die in a given year and the total size of each age group population in that same year. The method used here is complex and was programed by our colleagues in the California Department of Finance. There is more information about the specific methodology used on the technical tab.

"

ihmeTab <- paste0('

This (complex) chart shows (with its initial setting) estimates of which <b>factors</b> were associated with the most <b>Disability Adjusted Life Years</b> in California for two different years. Understanding which factors have the largest impact on health outcomes is essential for effective public health planning. Decreasing the level of detrimental risk factors (e.g. smoking) and increasing the level of healthy/protective factors (e.g. exercise) is the focus of most public health programs. By comparing the ranking of specific risk factors for each year observations can be made about changes in relative burden.

<br><br>

Constructing these estimates requires complicated procedures with many data inputs and many assumptions. The estimates here are based on sophisticated modeling by the Institute for Health Metrics and Evaluation (IHME) in Seattle. IHME generates such estimates for all US States, and for most nations of the world, but not yet for US counties. Information and resources related to these estimates can be found here.

<br><br>

In the initial setting for this chart, the health outcome being measured is Disability Adjusted Life Years (DALY). DALY is a powerful measure that sums (1) the number of years a person lives with a condition which decreases their functionality during life (Years Lived with Disability=YLD) and (2) the difference between the age at which they might have lived to without the condition age and the age at which they actually died due to the condition (Years of Life Lost=YLL); DALY = YLD + YLL. While DALYs is the initial health outcome setting, it can be changed to a different setting for the particular situation you are exploring.

The settings for all the modifiable settings (i.e. parameters), with their default values, options, and some explanations are:

<li style="margin-left: 20px">

Level of the risk factors</li>

<li style="margin-left: 40px">

Default: 2</li>

<li style="margin-left: 40px">

Options: 1, 2, 3</li>

<li style="margin-left: 40px">

These options are in increasing order of "granularity"; 1 is three very broad categories of risk factors (behavioral, metabolic, and behavioral); 2 breaks these down into 19 more detailed categories, and 3 into 25. </li>

<li style="margin-left: 20px">

Year: 1990 to 2017</li>

<li style="margin-left: 20px">

Sex: Female, Male, Total</li>

<li style="margin-left: 20px">

Metric for the Health Outcome: </li>

<li style="margin-left: 40px">

Default: Rate</li>

<li style="margin-left: 40px">

Options: Number, Percent, Rate</li>

<li style="margin-left: 20px">

Health Outcome</li>

<li style="margin-left: 40px">

Default: Disability Adjusted Life Years (DALYs) </li>

<li style="margin-left: 40px">

Options: Deaths, DALYS, Years Lived with Disability, Years of Life Lost</li>

')

hospA <-

"These charts show <b>rankings</b> for California and for each County of the <b>number of persons hospitalized</b> for specific reasons, the <b>total charges</b> for those hospitalizations, and the associated <b>median charges</b>. These rankings provide a valuable view into the burden of disease/injury in California, and provide an important view of the economic impact of these conditions.

<br><br>

The presentation of these hospitalization data should be considered preliminary. There are a number of nuances and considerations relevant to interpreting these charts noted in the technical documentation. Of particular note, the summaries of 'charges' shown, are specifically, <i>charges</i> based on hospital administrative systems, and do not indicate the actual costs, reimbursements, or payments for those charges. However, these charge data provide one lens of valuable information regarding the patterns of monetary burden. We are also in the midst of assessing optimal ways of grouping conditions/diseases and other aspects of the sharing of these data. We welcome your input."

hospB <-

"This chart shows the <b>primary</b> reason for hospitalization (i.e. the first code listed), and includes rankings based on the number of <b>hospitalizations</b>, the average <b>length of stay</b>, and associated <b>total charges</b> and <b>median charges</b>. This chart is particularly valuable for comparing the different rankings for the same condition based on numbers of hospitalizations, versus total charges or median charges. Some conditions have high (or low) total charges because of high (or low) median charges, some because of large (or small) numbers of hospitalizations, and all sorts of things in between."

hospC <-

"This chart shows hospitalization for a condition based on it being the primary reason (principal or primary diagnosis) for hospitalization OR it being listed in ANY of the other diagnostic positions (other or secondary diagnoses) for the hospitalization. This chart provides important insights for understanding burden since some conditions are overwhelmingly listed as 'primary' (e.g. birth-related), with few conditions listed in other positions; whereas other conditions are frequently listed in non-primary positions."

HospitalizationsTab <- paste(hospA,"<br><br>",hospB)

HospitalPrimaryAnyTab <- paste(hospA,"<br><br>",hospC)

disparitiesTab <- paste0('

This visualization is a new presentation of disparities data, and should be considered preliminary. We are exploring other approaches and methods and welcome your input.

<br><br>

This tab shows differences, or <b>disparities</b>, in death rates between <b>racial/ethnic groups</b>, <b>age groups</b>, and <b>sex</b>, for a selected cause of death in a selected county or the State overall. Reducing disparities in health outcomes is a key goal of Public Health, and clear data is essential for progress towards this goal.

<br><br>

<b>Within each demographic category</b> (race/ethnic, age, sex), for each cause of death in each geographic group, the <b>rate in the group with the <i>lowest rate</i> is statistically compared to each other group</b> in that category (e.g. for a condition where Hispanics have the lowest rate, that rate among Hispanics is compared to the rate among Blacks, Whites, and Asians, etc.).

<br><br>

In each chart, the group with the lowest rate is shown in green, any group that is statistically significantly higher (based on a statistical cut point of p < 0.01) is shown in red, and any group that is not statistically different from the "lowest" group is shown in blue. In addition, 95% confidence intervals are also shown for all rates, to provide visual guidance about the stability of each rate.

<br><br>

In many instances, not all groups within a demographic category will be shown because all rates based on fewer than 20 deaths are excluded, to avoid misleading observations from very unstable rates. ')