Project Proposal

Revised Proposal

Probability Pandas

10/17/2021

Load Packages

```
install.packages("tidyverse")
library(tidyverse)
```

Load Data

```
initial_data = read.csv("~/Probability Pandas Project/data/500_Cities__City-level_Data__GIS_Friendly_Fo
```

Introduction and Data, including Research Questions

(The introduction should introduce your general research question and your data (where it came from, how it was collected, what are the cases, what are the variables, etc.). Your research questions should be clearly specified. The motivation for your research question should be clear, with citations to relevant literature as appropriate.)

Glimpse

```
glimpse(initial_data)
```

```
## Rows: 500
## Columns: 117
## $ StateAbbr             <fct> CA, FL, CA, CA, FL, FL, NJ, CO, WI, WA, TX, IL,~
## $ PlaceName             <fct> Folsom, Largo, Berkeley, Napa, Sunrise, Pembrok~
## $ PlaceFIPS             <int> 624638, 1239425, 606000, 650258, 1269700, 12557~
## $ Population2010        <int> 72203, 77648, 112580, 76915, 84439, 154750, 849~
## $ ACCESS2_CrudePrev     <dbl> 7.5, 19.6, 7.7, 12.3, 22.8, 21.4, 25.4, 19.9, 1~
## $ ACCESS2_Crude95CI     <fct> "( 7.0,  8.0)", "(19.1, 20.2)", "( 7.3,  8.1)",~
## $ ACCESS2_AdjPrev       <dbl> 7.7, 20.9, 7.1, 12.7, 23.3, 22.0, 25.0, 20.6, 1~
## $ ACCESS2_Adj95CI       <fct> "( 7.2,  8.2)", "(20.4, 21.5)", "( 6.8,  7.3)",~
## $ ARTHRITIS_CrudePrev   <dbl> 16.9, 30.6, 15.1, 20.7, 22.8, 20.8, 23.9, 28.3,~
## $ ARTHRITIS_Crude95CI   <fct> "(16.6, 17.2)", "(30.3, 30.9)", "(15.0, 15.3)",~
## $ ARTHRITIS_AdjPrev     <dbl> 17.4, 23.3, 18.0, 19.3, 20.8, 18.8, 26.8, 25.7,~
## $ ARTHRITIS_Adj95CI     <fct> "(17.2, 17.7)", "(23.1, 23.5)", "(17.8, 18.1)",~
## $ BINGE_CrudePrev       <dbl> 21.8, 16.9, 19.6, 19.2, 16.3, 17.1, 14.4, 14.3,~
## $ BINGE_Crude95CI       <fct> "(21.5, 22.0)", "(16.8, 17.0)", "(19.4, 19.7)",~
## $ BINGE_AdjPrev         <dbl> 21.5, 20.6, 18.8, 19.9, 17.2, 18.2, 13.4, 15.2,~
```

```
## $ BINGE_Adj95CI          <fct> "(21.2, 21.7)", "(20.5, 20.8)", "(18.6, 18.9)",~
## $ BPHIGH_CrudePrev       <dbl> 25.7, 36.1, 20.9, 28.1, 33.3, 30.3, 35.4, 31.6,~
## $ BPHIGH_Crude95CI       <fct> "(25.3, 26.0)", "(35.8, 36.4)", "(20.8, 21.1)",~
## $ BPHIGH_AdjPrev         <dbl> 26.3, 28.4, 24.5, 26.5, 31.1, 27.9, 38.7, 29.0,~
## $ BPHIGH_Adj95CI         <fct> "(25.9, 26.6)", "(28.2, 28.7)", "(24.4, 24.7)",~
## $ BPMED_CrudePrev        <dbl> 64.8, 81.0, 68.2, 70.2, 76.7, 76.3, 74.4, 73.3,~
## $ BPMED_Crude95CI        <fct> "(64.3, 65.2)", "(80.9, 81.2)", "(68.0, 68.5)",~
## $ BPMED_AdjPrev          <dbl> 49.8, 58.4, 53.8, 50.8, 60.2, 57.9, 64.0, 53.5,~
## $ BPMED_Adj95CI          <fct> "(49.5, 50.2)", "(58.2, 58.6)", "(53.6, 54.0)",~
## $ CANCER_CrudePrev       <dbl> 5.8, 9.0, 4.9, 6.5, 6.5, 6.3, 4.6, 6.6, 5.9, 5.~
## $ CANCER_Crude95CI       <fct> "( 5.7,  5.8)", "( 8.9,  9.1)", "( 4.9,  4.9)",~
## $ CANCER_AdjPrev         <dbl> 6.2, 6.3, 6.0, 6.1, 5.8, 5.6, 5.4, 5.8, 6.0, 6.~
## $ CANCER_Adj95CI         <fct> "( 6.1,  6.3)", "( 6.3,  6.4)", "( 5.9,  6.0)",~
## $ CASTHMA_CrudePrev      <dbl> 8.6, 7.9, 8.8, 8.9, 8.0, 7.1, 10.8, 9.7, 10.1, ~
## $ CASTHMA_Crude95CI      <fct> "( 8.4,  8.7)", "( 7.8,  8.0)", "( 8.7,  8.9)",~
## $ CASTHMA_AdjPrev        <dbl> 8.5, 8.1, 8.7, 8.9, 8.0, 7.1, 10.7, 9.7, 10.1, ~
## $ CASTHMA_Adj95CI        <fct> "( 8.4,  8.7)", "( 8.0,  8.2)", "( 8.6,  8.7)",~
## $ CHD_CrudePrev          <dbl> 4.1, 9.8, 3.7, 5.8, 6.7, 6.1, 6.7, 7.2, 6.5, 5.~
## $ CHD_Crude95CI          <fct> "( 4.0,  4.2)", "( 9.6, 10.0)", "( 3.6,  3.7)",~
## $ CHD_AdjPrev            <dbl> 4.4, 6.7, 4.3, 5.3, 5.9, 5.3, 7.9, 6.2, 6.6, 6.~
## $ CHD_Adj95CI            <fct> "( 4.3,  4.6)", "( 6.5,  6.8)", "( 4.3,  4.4)",~
## $ CHECKUP_CrudePrev      <dbl> 64.7, 77.5, 64.7, 63.8, 77.7, 77.4, 76.7, 62.4,~
## $ CHECKUP_Crude95CI      <fct> "(64.4, 65.0)", "(77.4, 77.6)", "(64.5, 64.9)",~
## $ CHECKUP_AdjPrev        <dbl> 65.3, 73.6, 66.8, 62.8, 76.8, 76.3, 78.2, 60.6,~
## $ CHECKUP_Adj95CI        <fct> "(65.1, 65.6)", "(73.4, 73.8)", "(66.6, 67.0)",~
## $ CHOLSCREEN_CrudePrev   <dbl> 78.1, 80.2, 70.0, 75.4, 78.7, 80.6, 68.4, 70.9,~
## $ CHOLSCREEN_Crude95CI   <fct> "(77.5, 78.6)", "(79.9, 80.5)", "(69.6, 70.4)",~
## $ CHOLSCREEN_AdjPrev     <dbl> 77.3, 74.9, 77.8, 74.2, 77.2, 78.8, 71.2, 69.6,~
## $ CHOLSCREEN_Adj95CI     <fct> "(76.7, 77.9)", "(74.5, 75.3)", "(77.6, 78.1)",~
## $ COLON_SCREEN_CrudePrev <dbl> 76.6, 64.6, 75.4, 69.3, 59.7, 61.9, 53.9, 52.4,~
## $ COLON_SCREEN_Crude95CI <fct> "(75.7, 77.3)", "(63.9, 65.2)", "(75.0, 75.8)",~
## $ COLON_SCREEN_AdjPrev   <dbl> 77.7, 62.6, 74.6, 69.3, 59.9, 62.2, 55.0, 52.1,~
## $ COLON_SCREEN_Adj95CI   <fct> "(76.9, 78.4)", "(61.9, 63.3)", "(74.1, 75.0)",~
## $ COPD_CrudePrev         <dbl> 4.1, 10.0, 3.7, 5.9, 7.0, 5.7, 8.5, 6.9, 6.8, 6~
## $ COPD_Crude95CI         <fct> "( 3.9,  4.3)", "( 9.7, 10.3)", "( 3.6,  3.8)",~
## $ COPD_AdjPrev           <dbl> 4.2, 8.1, 4.1, 5.6, 6.5, 5.3, 9.3, 6.3, 6.8, 6.~
## $ COPD_Adj95CI           <fct> "( 4.0,  4.4)", "( 7.8,  8.3)", "( 4.0,  4.2)",~
## $ COREM_CrudePrev        <dbl> 37.1, 33.7, 38.2, 37.9, 30.5, 31.6, 23.0, 24.3,~
## $ COREM_Crude95CI        <fct> "(35.3, 39.0)", "(32.5, 34.9)", "(37.1, 39.3)",~
## $ COREM_AdjPrev          <dbl> 37.5, 33.9, 38.1, 38.3, 30.6, 31.2, 23.3, 24.6,~
## $ COREM_Adj95CI          <fct> "(35.4, 39.5)", "(32.7, 35.1)", "(37.2, 39.0)",~
## $ COREW_CrudePrev        <dbl> 33.3, 33.2, 36.6, 30.3, 26.2, 27.1, 20.9, 30.9,~
## $ COREW_Crude95CI        <fct> "(31.7, 35.1)", "(32.2, 34.4)", "(35.8, 37.4)",~
## $ COREW_AdjPrev          <dbl> 34.2, 34.4, 37.5, 31.4, 27.4, 28.0, 21.3, 31.9,~
## $ COREW_Adj95CI          <fct> "(32.5, 35.8)", "(33.3, 35.4)", "(36.8, 38.3)",~
## $ CSMOKING_CrudePrev     <dbl> 12.2, 20.7, 11.2, 14.5, 16.5, 13.1, 23.8, 19.1,~
## $ CSMOKING_Crude95CI     <fct> "(11.6, 12.8)", "(20.1, 21.1)", "(10.8, 11.7)",~
## $ CSMOKING_AdjPrev       <dbl> 11.7, 23.1, 11.4, 14.7, 16.9, 13.4, 23.1, 20.0,~
## $ CSMOKING_Adj95CI       <fct> "(11.1, 12.3)", "(22.5, 23.7)", "(11.1, 11.7)",~
## $ DENTAL_CrudePrev       <dbl> 74.7, 58.6, 70.0, 70.2, 61.0, 66.6, 52.8, 57.5,~
## $ DENTAL_Crude95CI       <fct> "(73.8, 75.5)", "(57.9, 59.4)", "(69.3, 70.7)",~
## $ DENTAL_AdjPrev         <dbl> 74.9, 57.6, 71.7, 70.0, 60.8, 66.4, 53.3, 57.0,~
## $ DENTAL_Adj95CI         <fct> "(74.0, 75.7)", "(56.8, 58.3)", "(71.1, 72.2)",~
## $ DIABETES_CrudePrev     <dbl> 6.7, 12.1, 6.5, 8.9, 12.1, 11.1, 13.7, 11.0, 11~
```

```
## $ DIABETES_Crude95CI     <fct> "( 6.6,  6.9)", "(11.9, 12.3)", "( 6.4,  6.6)",~
## $ DIABETES_AdjPrev       <dbl> 6.9, 9.2, 7.7, 8.3, 11.2, 10.1, 15.4, 10.0, 11.~
## $ DIABETES_Adj95CI       <fct> "( 6.7,  7.0)", "( 9.1,  9.3)", "( 7.6,  7.8)",~
## $ HIGHCHOL_CrudePrev     <dbl> 29.1, 39.0, 27.1, 34.1, 37.1, 36.1, 36.0, 37.6,~
## $ HIGHCHOL_Crude95CI     <fct> "(28.8, 29.4)", "(38.7, 39.2)", "(26.9, 27.3)",~
## $ HIGHCHOL_AdjPrev       <dbl> 26.7, 29.8, 26.4, 29.2, 32.1, 31.2, 34.1, 31.1,~
## $ HIGHCHOL_Adj95CI       <fct> "(26.4, 26.9)", "(29.6, 30.0)", "(26.3, 26.6)",~
## $ KIDNEY_CrudePrev       <dbl> 2.1, 3.7, 2.1, 2.8, 3.2, 3.0, 3.5, 3.6, 3.1, 2.~
## $ KIDNEY_Crude95CI       <fct> "( 2.1,  2.1)", "( 3.7,  3.8)", "( 2.1,  2.1)",~
## $ KIDNEY_AdjPrev         <dbl> 2.2, 2.8, 2.4, 2.6, 3.0, 2.7, 4.0, 3.3, 3.2, 3.~
## $ KIDNEY_Adj95CI         <fct> "( 2.1,  2.2)", "( 2.8,  2.9)", "( 2.4,  2.4)",~
## $ LPA_CrudePrev          <dbl> 14.3, 31.0, 14.2, 19.8, 29.5, 26.4, 42.2, 24.2,~
## $ LPA_Crude95CI          <fct> "(13.8, 14.8)", "(30.4, 31.6)", "(13.8, 14.6)",~
## $ LPA_AdjPrev            <dbl> 14.4, 28.6, 14.5, 19.5, 28.6, 25.5, 43.4, 23.6,~
## $ LPA_Adj95CI            <fct> "(13.8, 14.9)", "(28.1, 29.2)", "(14.2, 14.8)",~
## $ MAMMOUSE_CrudePrev     <dbl> 80.4, 75.7, 81.5, 76.7, 82.5, 82.7, 79.7, 72.9,~
## $ MAMMOUSE_Crude95CI     <fct> "(79.6, 81.2)", "(75.0, 76.3)", "(81.0, 81.9)",~
## $ MAMMOUSE_AdjPrev       <dbl> 78.8, 71.3, 78.2, 73.4, 79.0, 80.3, 72.6, 66.2,~
## $ MAMMOUSE_Adj95CI       <fct> "(77.9, 79.7)", "(70.6, 72.0)", "(77.7, 78.7)",~
## $ MHLTH_CrudePrev        <dbl> 9.9, 13.1, 10.9, 12.0, 12.7, 10.9, 17.0, 13.5, ~
## $ MHLTH_Crude95CI        <fct> "( 9.6, 10.2)", "(12.9, 13.4)", "(10.6, 11.2)",~
## $ MHLTH_AdjPrev          <dbl> 9.7, 14.4, 10.1, 12.1, 13.0, 11.1, 16.5, 13.8, ~
## $ MHLTH_Adj95CI          <fct> "( 9.4,  9.9)", "(14.1, 14.6)", "( 9.9, 10.3)",~
## $ OBESITY_CrudePrev      <dbl> 23.8, 28.3, 18.5, 24.0, 28.1, 25.7, 38.9, 30.0,~
## $ OBESITY_Crude95CI      <fct> "(23.5, 24.1)", "(28.1, 28.6)", "(18.3, 18.7)",~
## $ OBESITY_AdjPrev        <dbl> 23.1, 28.8, 20.6, 24.0, 28.2, 25.6, 39.3, 30.7,~
## $ OBESITY_Adj95CI        <fct> "(22.7, 23.4)", "(28.6, 29.1)", "(20.5, 20.8)",~
## $ PAPTEST_CrudePrev      <dbl> 84.3, 77.1, 83.2, 83.9, 81.3, 82.2, 84.9, 81.3,~
## $ PAPTEST_Crude95CI      <fct> "(83.7, 84.9)", "(76.6, 77.7)", "(82.7, 83.7)",~
## $ PAPTEST_AdjPrev        <dbl> 82.9, 71.5, 81.3, 81.1, 77.0, 79.6, 77.4, 74.3,~
## $ PAPTEST_Adj95CI        <fct> "(82.3, 83.5)", "(70.9, 72.2)", "(80.9, 81.7)",~
## $ PHLTH_CrudePrev        <dbl> 8.9, 15.4, 8.2, 12.0, 13.3, 11.6, 17.8, 16.1, 1~
## $ PHLTH_Crude95CI        <fct> "( 8.6,  9.2)", "(15.0, 15.7)", "( 8.0,  8.3)",~
## $ PHLTH_AdjPrev          <dbl> 8.8, 13.6, 9.0, 11.6, 12.7, 10.9, 18.7, 15.5, 1~
## $ PHLTH_Adj95CI          <fct> "( 8.5,  9.1)", "(13.3, 13.8)", "( 8.8,  9.1)",~
## $ SLEEP_CrudePrev        <dbl> 33.9, 37.7, 32.2, 32.7, 38.1, 35.8, 45.7, 34.2,~
## $ SLEEP_Crude95CI        <fct> "(33.5, 34.3)", "(37.4, 38.0)", "(31.9, 32.5)",~
## $ SLEEP_AdjPrev          <dbl> 33.1, 39.8, 32.5, 33.0, 38.5, 36.1, 45.2, 35.1,~
## $ SLEEP_Adj95CI          <fct> "(32.7, 33.5)", "(39.5, 40.1)", "(32.3, 32.7)",~
## $ STROKE_CrudePrev       <dbl> 1.9, 4.5, 1.9, 2.8, 3.7, 3.1, 4.4, 3.8, 3.5, 3.~
## $ STROKE_Crude95CI       <fct> "( 1.9,  2.0)", "( 4.4,  4.7)", "( 1.8,  1.9)",~
## $ STROKE_AdjPrev         <dbl> 2.0, 3.2, 2.3, 2.6, 3.3, 2.7, 5.1, 3.3, 3.6, 3.~
## $ STROKE_Adj95CI         <fct> "( 2.0,  2.1)", "( 3.2,  3.3)", "( 2.2,  2.3)",~
## $ TEETHLOST_CrudePrev    <dbl> 6.8, 18.3, 6.7, 11.2, 16.2, 14.1, 26.1, 17.7, 1~
## $ TEETHLOST_Crude95CI    <fct> "( 5.7,  8.0)", "(16.9, 19.8)", "( 6.2,  7.2)",~
## $ TEETHLOST_AdjPrev      <dbl> 6.8, 18.0, 6.8, 11.2, 15.8, 13.6, 26.4, 17.4, 1~
## $ TEETHLOST_Adj95CI      <fct> "( 5.7,  7.9)", "(16.7, 19.4)", "( 6.3,  7.4)",~
## $ Geolocation            <fct> "(38.67504943280, -121.147605753)", "(27.909090~
```

Data Analysis Plan

(Specify the outcome (response, Y) and predictor (explanatory, X) variables you will use to answer your question, as well as the comparison groups you will use, if applicable. You may include very preliminary

exploratory data analysis, including some summary statistics and visualizations, along with some explanation on how they help you learn more about your data. Note the statistical method(s) that you believe will be useful in answering your question(s). What results from these specific statistical methods are needed to support your hypothesized answer?)