

R Project Milestone 2

Moyra Rasheed, Courtney Coon, Jarett Maycott

2022-09-30

Importing data

demographic dataset

```
demo_data<-read.csv(  
  "https://raw.githubusercontent.com/PHW290/phw251_projectdata/main/ca_county_demographic.csv",  
  stringsAsFactors = FALSE)  
str(demo_data)
```

```
## 'data.frame':   58 obs. of  22 variables:  
## $ name       : chr  "Kern" "Kings" "Lake" "Lassen" ...  
## $ pop2012    : int   851089 155039 65253 35039 9904341 153025 255509 18455 88094 256841 ...  
## $ pop12_sqmi : num   104.28 111.43 49.08 7.42 2423.26 ...  
## $ white      : int   499766 83027 52033 25532 4936599 94456 201963 16103 67218 148381 ...  
## $ black      : int   48921 11014 1232 2834 856874 5629 6987 138 622 9926 ...  
## $ ameri_es   : int   12676 2562 2049 1234 72828 4136 1523 527 4277 3473 ...  
## $ asian      : int   34846 5620 724 356 1346865 2802 13761 204 1450 18836 ...  
## $ hawn_pi    : int   1252 271 108 165 26094 162 509 26 119 583 ...  
## $ hispanic   : int   413033 77866 11088 6117 4687889 80992 39069 1676 19505 140485 ...  
## $ other      : int   204314 42996 5455 3562 2140632 37380 16973 508 10185 62665 ...  
## $ mult_race  : int   37856 7492 3064 1212 438713 6300 10693 745 3970 11929 ...  
## $ males      : int   433108 86344 32469 22416 4839654 72682 124072 9269 43983 128737 ...  
## $ females    : int   406523 66638 32196 12479 4978951 78183 128337 8982 43858 127056 ...  
## $ med_age    : num    30.7 31.1 45 37 34.8 33.1 44.5 49.2 41.6 29.6 ...  
## $ households : int   254610 41233 26548 10058 3241204 43317 103210 7693 34945 75642 ...  
## $ families   : int   191739 31939 16255 6800 2194080 34093 62653 4948 21591 58767 ...  
## $ hse_units  : int   284367 43867 35492 12710 3445076 49140 111214 10188 40323 83698 ...  
## $ ave_fam_sz : num     3.61 3.59 2.94 2.98 3.58 3.63 2.94 2.77 3.02 3.74 ...  
## $ vacant     : int   29757 2634 8944 2652 203872 5823 8004 2495 5378 8056 ...  
## $ owner_occ  : int   152828 22329 17472 6590 1544749 27726 64637 5227 20601 41196 ...  
## $ renter_occ : int   101782 18904 9076 3468 1696455 15591 38573 2466 14344 34446 ...  
## $ county_fips: int    6103 6089 6106 6086 6073 6102 6066 6111 6100 6099 ...
```

```
#data is clean and ready to use
```

mortality dataset

```
mort_data<-read.csv(
  "https://raw.githubusercontent.com/PHW290/phw251_projectdata/main/ca_county_mortality.csv",
  stringsAsFactors = FALSE, na.strings = "")

#remove last 2 columns b/c mostly NAs ("Annotation...")
mort_data<-mort_data[,1:8]
head(mort_data)
```

```
##   Year County Geography_Type      Strata      Strata_Name Cause
## 1 2014 Alameda      Occurrence Total Population Total Population  ALL
## 2 2014 Alameda      Occurrence      Age      Under 1 year  ALL
## 3 2014 Alameda      Occurrence      Age        1-4 years  ALL
## 4 2014 Alameda      Occurrence      Age        5-14 years  ALL
## 5 2014 Alameda      Occurrence      Age       15-24 years  ALL
## 6 2014 Alameda      Occurrence      Age       25-34 years  ALL
##           Cause_Desc Count
## 1 All causes (total)  9357
## 2 All causes (total)   105
## 3 All causes (total)    17
## 4 All causes (total)    17
## 5 All causes (total)   133
## 6 All causes (total)   175
```

```
#column names are all capitalized
colnames(mort_data) <- str_to_lower(colnames(mort_data))
head(mort_data)
```

```
##   year  county geography_type      strata      strata_name cause
## 1 2014 Alameda      Occurrence Total Population Total Population  ALL
## 2 2014 Alameda      Occurrence      Age      Under 1 year  ALL
## 3 2014 Alameda      Occurrence      Age        1-4 years  ALL
## 4 2014 Alameda      Occurrence      Age        5-14 years  ALL
## 5 2014 Alameda      Occurrence      Age       15-24 years  ALL
## 6 2014 Alameda      Occurrence      Age       25-34 years  ALL
##           cause_desc count
## 1 All causes (total)  9357
## 2 All causes (total)   105
## 3 All causes (total)    17
## 4 All causes (total)    17
## 5 All causes (total)   133
## 6 All causes (total)   175
```

mortality dataset con't

```
#interested in knowing categories of mortality  
unique(mort_data$cause_desc)
```

```
## [1] "All causes (total)"  
## [2] "Alzheimer's disease"  
## [3] "Malignant neoplasms"  
## [4] "Chronic lower respiratory diseases"  
## [5] "Diabetes mellitus"  
## [6] "Assault (homicide)"  
## [7] "Diseases of heart"  
## [8] "Essential hypertension and hypertensive renal disease"  
## [9] "Accidents (unintentional injuries)"  
## [10] "Chronic liver disease and cirrhosis"  
## [11] "Nephritis, nephrotic syndrome and nephrosis"  
## [12] "Parkinson's disease"  
## [13] "Influenza and pneumonia"  
## [14] "Cerebrovascular diseases"  
## [15] "Intentional self-harm (suicide)"
```

```
unique(mort_data$year)
```

```
## [1] 2014 2015 2016 2017 2018 2019 2020
```

```
#note "strata" and "strata_name" may have categories we want to pull out at some point
```

HCAI Healthcare dataset

```
healthcare_data<-read.csv(  
  "https://raw.githubusercontent.com/PHW290/phw251_projectdata/main/hcai_healthcare_construction.csv",  
  stringsAsFactors = FALSE,na.strings = "")  
  
#remove last column ("Collection.of.Counties") b/c mostly NAs  
healthcare_data<-healthcare_data[,1:5]  
head(healthcare_data)
```

```
##      County Data.Generation.Date OSHPD.Project.Status  
## 1 01 - Alameda 2013-10-14T00:00:00      In Review  
## 2 01 - Alameda 2013-10-14T00:00:00 Pending Construction  
## 3 01 - Alameda 2013-10-14T00:00:00      In Construction  
## 4 01 - Alameda 2013-10-14T00:00:00      In Closure  
## 5 02 - Alpine 2013-10-14T00:00:00      In Review  
## 6 02 - Alpine 2013-10-14T00:00:00 Pending Construction  
##      Total.Costs.of.OSHPD.Projects Number.of.OSHPD.Projects  
## 1                $50,890,315.00                44  
## 2                $840,242,543.36                125  
## 3                $994,245,713.95                181  
## 4                $65,337,613.88                 82  
## 5                  $0.00                      0  
## 6                  $0.00                      0
```

```
#column names are all capitalized  
colnames(healthcare_data) <- str_to_lower(colnames(healthcare_data))  
head(healthcare_data)
```

```
##      county data.generation.date oshpd.project.status  
## 1 01 - Alameda 2013-10-14T00:00:00      In Review  
## 2 01 - Alameda 2013-10-14T00:00:00 Pending Construction  
## 3 01 - Alameda 2013-10-14T00:00:00      In Construction  
## 4 01 - Alameda 2013-10-14T00:00:00      In Closure  
## 5 02 - Alpine 2013-10-14T00:00:00      In Review  
## 6 02 - Alpine 2013-10-14T00:00:00 Pending Construction  
##      total.costs.of.oshpd.projects number.of.oshpd.projects  
## 1                $50,890,315.00                44  
## 2                $840,242,543.36                125  
## 3                $994,245,713.95                181  
## 4                $65,337,613.88                 82  
## 5                  $0.00                      0  
## 6                  $0.00                      0
```

HCAI Healthcare dataset con't

```
#change "." to "_" for consistency with other datasets
names(healthcare_data) <- gsub(x = names(healthcare_data), pattern = "\\.",
                               replacement = "_")
head(healthcare_data)
```

```
##      county data_generation_date oshpd_project_status
## 1 01 - Alameda 2013-10-14T00:00:00      In Review
## 2 01 - Alameda 2013-10-14T00:00:00 Pending Construction
## 3 01 - Alameda 2013-10-14T00:00:00      In Construction
## 4 01 - Alameda 2013-10-14T00:00:00      In Closure
## 5 02 - Alpine 2013-10-14T00:00:00      In Review
## 6 02 - Alpine 2013-10-14T00:00:00 Pending Construction
## total_costs_of_oshpd_projects number_of_oshpd_projects
## 1                $50,890,315.00                44
## 2                $840,242,543.36                125
## 3                $994,245,713.95                181
## 4                $65,337,613.88                 82
## 5                  $0.00                      0
## 6                  $0.00                      0
```

```
#change county names to match other two datasets (remove numbers in front)
healthcare_data<-healthcare_data%>%
  mutate(county=substring(healthcare_data$county, 6))
head(healthcare_data)
```

```
##      county data_generation_date oshpd_project_status
## 1 Alameda 2013-10-14T00:00:00      In Review
## 2 Alameda 2013-10-14T00:00:00 Pending Construction
## 3 Alameda 2013-10-14T00:00:00      In Construction
## 4 Alameda 2013-10-14T00:00:00      In Closure
## 5 Alpine 2013-10-14T00:00:00      In Review
## 6 Alpine 2013-10-14T00:00:00 Pending Construction
## total_costs_of_oshpd_projects number_of_oshpd_projects
## 1                $50,890,315.00                44
## 2                $840,242,543.36                125
## 3                $994,245,713.95                181
## 4                $65,337,613.88                 82
## 5                  $0.00                      0
## 6                  $0.00                      0
```

HCAI Healthcare dataset con't

```
#change money column ("Total.Costs...") to numeric
healthcare_data<-healthcare_data%>%
  mutate(total_costs_of_oshpd_projects = total_costs_of_oshpd_projects %>%
    str_remove_all("[$,]"))
healthcare_data$total_costs_of_oshpd_projects<-as.numeric(healthcare_data$total_costs_of_oshpd_projects)
head(healthcare_data)
```

```
##      county data_generation_date oshpd_project_status
## 1 Alameda 2013-10-14T00:00:00      In Review
## 2 Alameda 2013-10-14T00:00:00 Pending Construction
## 3 Alameda 2013-10-14T00:00:00      In Construction
## 4 Alameda 2013-10-14T00:00:00      In Closure
## 5 Alpine 2013-10-14T00:00:00      In Review
## 6 Alpine 2013-10-14T00:00:00 Pending Construction
##      total_costs_of_oshpd_projects number_of_oshpd_projects
## 1                      50890315                      44
## 2                      840242543                     125
## 3                      994245714                     181
## 4                      65337614                      82
## 5                          0                          0
## 6                          0                          0
```

```
#fix "data_generation_date" (remove empty time)
healthcare_data$data_generation_date<-as.Date(healthcare_data$data_generation_date)
head(healthcare_data)
```

```
##      county data_generation_date oshpd_project_status
## 1 Alameda      2013-10-14      In Review
## 2 Alameda      2013-10-14 Pending Construction
## 3 Alameda      2013-10-14      In Construction
## 4 Alameda      2013-10-14      In Closure
## 5 Alpine       2013-10-14      In Review
## 6 Alpine       2013-10-14 Pending Construction
##      total_costs_of_oshpd_projects number_of_oshpd_projects
## 1                      50890315                      44
## 2                      840242543                     125
## 3                      994245714                     181
## 4                      65337614                      82
## 5                          0                          0
## 6                          0                          0
```

Description of dataset

All 3 data sets are CSV files.

What is the data source? How does the dataset relate to the group problem statement and question?

1. Demographic: Census Data

Includes 58 observations with 22 variables. All are numeric or integers except for the first column which is a character with county names.

2. Mortality: CA open data portal, California Department of Public Health

Includes 147,784 observations with 10 variables. most are characters except for year and count of mortality.

3. Healthcare: CA open data portal, Department of Healthcare Access and Information

Includes 53,592 observations with 5 variables.

How does the dataset relate to the group problem statement and question? Allocated funding for healthcare for vulnerable populations in California. Certain regions in California are considered to be vulnerable populations. Given that they need access to better funding, thereby providing access to better healthcare, we are able to allocate funds to certain regions. By utilizing all 3 data sets, we can focus on rural populations, seniors and renters, higher mortality rate and the counties that have average or below average funding available in the last 5 years.

Identify data types for 5+ data elements/columns/variables

Identify 5+ data elements required for your specified scenario.

If <5 elements are required to complete the analysis, please choose additional variables of interest in the data set to explore in this milestone.

1. Demographic Data: demo_data\$pop12_sqmi: Type:chr
2. Demographic Data: demo_data\$name: Type:chr
3. Demographic: demo_data\$med_age: Type:num
4. Demographics: demo_data\$renter_occ: Type:num
5. Demographics: demo_data\$owner_occ: Type:num
6. Mortality: mort_data\$county: Type:chr
7. Mortality: mort_data\$year: Type:chr
8. Mortality: mort_data\$cause_desc: Type: chr
9. Mortality: mort_data\$count: Type:num
10. HCAI healthcare: healthcare_data\$oshph_project_status: Type:chr
11. HCAI healthcare: healthcare_data\$data_generation_date: Type:chr
12. HCAI healthcare: healthcare_data\$total_costs_of_oshpd_projects: Type:num
13. HCAI healthcare: healthcare_data\$county: Type:chr
14. HCAI healthcare: healthcare_data\$number_of_oshpd_projects: Type:num

Identify the desired type/format for each variable—will you need to convert any columns to numeric or another type?

1. HCAI: data_generation_date: change format from date-time to strictly date;

may need to change to year to merge data for Milestone #3

2. HCAI: changed “total_cost_projects” from chr (string) to numeric by removing “\$”

The remaining columns are in the appropriate format for now.

Provide a basic description of the 5+ data elements (Numeric: mean, median, range; Character: unique values/categories)

```
#1: demo_data$pop12_sqmi
summary(demo_data$pop12_sqmi)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
##    1.544    25.887   103.424   665.061   333.485 17398.354
```

```
#2: demo_data$name
unique(demo_data$name)
```

```
## [1] "Kern"           "Kings"           "Lake"            "Lassen"
## [5] "Los Angeles"    "Madera"          "Marin"           "Mariposa"
## [9] "Mendocino"      "Merced"          "Modoc"           "Mono"
## [13] "Monterey"       "Napa"            "Nevada"          "Orange"
## [17] "Placer"         "Plumas"          "Riverside"       "Sacramento"
## [21] "San Benito"     "San Bernardino" "San Diego"       "San Francisco"
## [25] "San Joaquin"    "San Luis Obispo" "San Mateo"       "Santa Barbara"
## [29] "Santa Clara"    "Santa Cruz"      "Shasta"          "Sierra"
## [33] "Siskiyou"       "Solano"          "Alameda"         "Alpine"
## [37] "Sonoma"         "Amador"          "Stanislaus"      "Sutter"
## [41] "Butte"          "Calaveras"       "Tehama"          "Colusa"
## [45] "Trinity"        "Tulare"          "Contra Costa"    "Del Norte"
## [49] "Tuolumne"       "Ventura"         "El Dorado"       "Yolo"
## [53] "Fresno"         "Glenn"           "Yuba"            "Humboldt"
## [57] "Imperial"       "Inyo"
```

```
#3: demo_data$name
summary(demo_data$name)
```

```
##      Length      Class      Mode
##          58 character character
```

```
#4: demo_data$renter_occ
summary(demo_data$renter_occ)
```

```
##      Min. 1st Qu.  Median     Mean 3rd Qu.     Max.
##     140   6080   25140   95554   84189 1696455
```

```
#5: demo_data$owner_occ
summary(demo_data$owner_occ)
```

```
##      Min. 1st Qu.  Median     Mean 3rd Qu.     Max.
##     357  13089   39306  121300  120804 1544749
```

```
#6: mort_data$county
unique(mort_data$county)
```

```
## [1] "Alameda"      "Alpine"      "Amador"      "Butte"
## [5] "Calaveras"    "Colusa"      "Contra Costa" "Del Norte"
## [9] "El Dorado"    "Fresno"      "Glenn"        "Humboldt"
## [13] "Imperial"     "Inyo"        "Kern"         "Kings"
## [17] "Lake"         "Lassen"      "Los Angeles"  "Madera"
## [21] "Marin"        "Mariposa"    "Mendocino"    "Merced"
## [25] "Modoc"        "Mono"        "Monterey"     "Napa"
## [29] "Nevada"       "Orange"      "Placer"       "Plumas"
## [33] "Riverside"    "Sacramento"  "San Benito"   "San Bernardino"
## [37] "San Diego"    "San Francisco" "San Joaquin"  "San Luis Obispo"
## [41] "San Mateo"    "Santa Barbara" "Santa Clara"  "Santa Cruz"
## [45] "Shasta"       "Sierra"      "Siskiyou"     "Solano"
## [49] "Sonoma"       "Stanislaus"  "Sutter"       "Tehama"
## [53] "Trinity"      "Tulare"      "Tuolumne"     "Ventura"
## [57] "Yolo"         "Yuba"
```

```
#7: mort_data$year
unique(mort_data$year)
```

```
## [1] 2014 2015 2016 2017 2018 2019 2020
```

```
#8: mort_data$cause_desc
unique(mort_data$cause_desc)
```

```
## [1] "All causes (total)"
## [2] "Alzheimer's disease"
## [3] "Malignant neoplasms"
## [4] "Chronic lower respiratory diseases"
## [5] "Diabetes mellitus"
## [6] "Assault (homicide)"
## [7] "Diseases of heart"
## [8] "Essential hypertension and hypertensive renal disease"
## [9] "Accidents (unintentional injuries)"
## [10] "Chronic liver disease and cirrhosis"
## [11] "Nephritis, nephrotic syndrome and nephrosis"
## [12] "Parkinson's disease"
## [13] "Influenza and pneumonia"
## [14] "Cerebrovascular diseases"
## [15] "Intentional self-harm (suicide)"
```

```
#9: mort_data$count
summary(mort_data$count)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      0.0     0.0    15.0   273.8   94.0 82816.0  45348
```

```
#10: healthcare_data$oshpd_project_status
unique(healthcare_data$oshpd_project_status)
```

```
## [1] "In Review" "Pending Construction" "In Construction"
## [4] "In Closure"
```

```
#11: healthcare_data$data_generation_date
unique(healthcare_data$data_generation_date)
```

```
## [1] "2013-10-14" "2013-10-15" "2013-10-31" "2013-11-14" "2013-11-26"
## [6] "2013-12-12" "2014-01-02" "2014-01-21" "2014-01-30" "2014-02-14"
## [11] "2014-03-03" "2014-03-19" "2014-04-07" "2014-04-21" "2014-05-08"
## [16] "2014-05-27" "2014-06-13" "2014-06-23" "2014-07-03" "2014-07-17"
## [21] "2014-07-30" "2014-08-14" "2014-08-28" "2014-09-19" "2014-09-24"
## [26] "2014-10-08" "2014-10-23" "2014-11-07" "2014-11-21" "2014-12-01"
## [31] "2014-12-15" "2014-12-23" "2015-01-05" "2015-01-15" "2015-01-29"
## [36] "2015-02-13" "2015-02-26" "2015-03-13" "2015-03-27" "2015-04-09"
## [41] "2015-04-23" "2015-05-07" "2015-05-14" "2015-05-21" "2015-05-28"
## [46] "2015-06-04" "2015-06-23" "2015-07-02" "2015-07-20" "2015-07-30"
## [51] "2015-08-12" "2015-08-27" "2015-09-09" "2015-09-24" "2015-10-08"
## [56] "2015-10-22" "2015-11-05" "2015-11-19" "2015-12-03" "2015-12-17"
## [61] "2015-12-31" "2016-01-14" "2016-01-28" "2016-02-11" "2016-02-29"
## [66] "2016-03-10" "2016-03-24" "2016-04-07" "2016-04-21" "2016-05-05"
## [71] "2016-05-19" "2016-06-01" "2016-06-16" "2016-06-30" "2016-07-11"
## [76] "2016-07-28" "2016-08-11" "2016-08-25" "2016-09-08" "2016-09-22"
## [81] "2016-10-07" "2016-10-20" "2016-11-03" "2016-11-17" "2016-12-01"
## [86] "2016-12-15" "2016-12-29" "2017-01-12" "2017-01-26" "2017-02-09"
## [91] "2017-02-23" "2017-03-09" "2017-03-23" "2017-04-06" "2017-04-20"
## [96] "2017-05-04" "2017-05-18" "2017-06-01" "2017-06-15" "2017-06-29"
## [101] "2017-07-17" "2017-07-28" "2017-08-10" "2017-08-25" "2017-09-08"
## [106] "2017-09-21" "2017-10-06" "2017-10-20" "2017-11-02" "2017-11-17"
## [111] "2017-11-30" "2017-12-15" "2017-12-28" "2018-01-12" "2018-01-25"
## [116] "2018-02-08" "2018-02-23" "2018-03-08" "2018-03-22" "2018-04-06"
## [121] "2018-04-19" "2018-05-04" "2018-05-17" "2018-05-31" "2018-06-14"
## [126] "2018-06-28" "2018-07-16" "2018-07-26" "2018-08-09" "2018-08-23"
## [131] "2018-09-06" "2018-09-20" "2018-10-04" "2018-10-18" "2018-11-01"
## [136] "2018-11-15" "2018-11-30" "2018-12-13" "2018-12-27" "2019-01-10"
## [141] "2019-01-24" "2019-02-07" "2019-02-22" "2019-03-07" "2019-03-21"
## [146] "2019-04-05" "2019-04-19" "2019-05-02" "2019-05-16" "2019-05-30"
## [151] "2019-06-13" "2019-06-27" "2019-07-11" "2019-07-24" "2019-08-08"
## [156] "2019-08-26" "2019-09-05" "2019-09-19" "2019-10-03" "2019-10-17"
## [161] "2019-10-31" "2019-11-14" "2019-11-26" "2019-12-13" "2019-12-24"
## [166] "2020-01-09" "2020-01-23" "2020-02-05" "2020-02-20" "2020-03-05"
## [171] "2020-04-02" "2020-04-10" "2020-04-23" "2020-05-07" "2020-05-21"
## [176] "2020-06-04" "2020-06-18" "2020-07-02" "2020-07-17" "2020-07-30"
## [181] "2020-08-14" "2020-08-27" "2020-09-10" "2020-09-25" "2020-10-08"
## [186] "2020-10-23" "2020-11-06" "2020-11-19" "2020-12-04" "2020-12-18"
## [191] "2020-12-30" "2021-01-14" "2021-01-28" "2021-02-11" "2021-02-25"
## [196] "2021-03-11" "2021-03-25" "2021-04-09" "2021-05-06" "2021-05-21"
## [201] "2021-06-03" "2021-06-17" "2021-07-01" "2021-07-15" "2021-08-02"
## [206] "2021-08-12" "2021-08-26" "2021-09-09" "2021-09-23" "2021-10-07"
## [211] "2021-10-21" "2021-11-05" "2021-11-18" "2021-12-02" "2021-12-16"
```

```
## [216] "2022-01-06" "2022-01-27" "2022-02-10" "2022-02-24" "2022-03-10"
## [221] "2022-03-24" "2022-04-07" "2022-04-21" "2022-05-05" "2022-05-19"
## [226] "2022-06-02" "2022-06-16" "2022-06-30" "2022-07-18" "2022-07-28"
## [231] "2022-08-11"
```

```
#12: healthcare_data$total_costs_of_oshpd_projects
summary(healthcare_data$total_costs_of_oshpd_projects)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## 0.000e+00 9.807e+04 2.824e+06 5.914e+07 2.845e+07 2.340e+09
```

```
#13: healthcare_data$county
unique(healthcare_data$county)
```

```
## [1] "Alameda"      "Alpine"      "Amador"      "Butte"
## [5] "Calaveras"    "Colusa"      "Contra Costa" "Del Norte"
## [9] "El Dorado"    "Fresno"      "Glenn"        "Humboldt"
## [13] "Imperial"     "Inyo"        "Kern"         "Kings"
## [17] "Lake"         "Lassen"      "Los Angeles"  "Madera"
## [21] "Marin"        "Mariposa"    "Mendocino"    "Merced"
## [25] "Modoc"        "Mono"        "Monterey"     "Napa"
## [29] "Nevada"       "Orange"      "Placer"       "Plumas"
## [33] "Riverside"    "Sacramento"  "San Benito"   "San Bernardino"
## [37] "San Diego"    "San Francisco" "San Joaquin"  "San Luis Obispo"
## [41] "San Mateo"    "Santa Barbara" "Santa Clara"  "Santa Cruz"
## [45] "Shasta"      "Sierra"      "Siskiyou"     "Solano"
## [49] "Sonoma"      "Stanislaus"  "Sutter"       "Tehama"
## [53] "Trinity"     "Tulare"     "Tuolumne"     "Ventura"
## [57] "Yolo"        "Yuba"
```

```
#14: healthcare_data$number_of_oshpd_projects
summary(healthcare_data$number_of_oshpd_projects)
```

```
##      Min. 1st Qu.  Median     Mean 3rd Qu.     Max.
##      0.00   1.00    6.00   27.94   23.00  1055.00
```