DSC520 Week 02 Assignment

Kurt Stoneburner

6/11/2020

What are the elements in your data (including the categories and data types)

1. There are 8 elements in the dataset

- Id : Categorical, used as an identifier
- Id2 : Categorical, used as an identifier
- Geography : Categorical
- PopGroupID : Categorical
- RacesReported : Quantitative
- HSDegree : Quantitative
- BachDegree : Quantitative

2. **head(survey_df)**

```
##               Id  Id2                         Geography PopGroupID
## 1 0500000US01073 1073       Jefferson County, Alabama          1
## 2 0500000US04013 4013        Maricopa County, Arizona          1
## 3 0500000US04019 4019            Pima County, Arizona          1
## 4 0500000US06001 6001      Alameda County, California          1
## 5 0500000US06013 6013 Contra Costa County, California          1
## 6 0500000US06019 6019       Fresno County, California          1
##   POPGROUP.display.label RacesReported HSDegree BachDegree
## 1       Total population        660793     89.1       30.5
## 2       Total population       4087191     86.8       30.2
## 3       Total population       1004516     88.0       30.8
## 4       Total population       1610921     86.9       42.8
## 5       Total population       1111339     88.8       39.7
## 6       Total population        965974     73.6       19.7
```

str(survey_df)

```
## 'data.frame':    136 obs. of  8 variables:
##  $ Id                   : chr  "0500000US01073" "0500000US04013" "0500000US04019" "0500000US06001"
##  $ Id2                  : int  1073 4013 4019 6001 6013 6019 6029 6037 6059 6065 ...
##  $ Geography            : chr  "Jefferson County, Alabama" "Maricopa County, Arizona" "Pima County,
##  $ PopGroupID           : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ POPGROUP.display.label: chr  "Total population" "Total population" "Total population" "Total popul
##  $ RacesReported        : int  660793 4087191 1004516 1610921 1111339 965974 874589 10116705 314551
##  $ HSDegree             : num  89.1 86.8 88 86.9 88.8 73.6 74.5 77.5 84.6 80.6 ...
##  $ BachDegree           : num  30.5 30.2 30.8 42.8 39.7 19.7 15.4 30.3 38 20.7 ...
```

nrow(survey_df)

```
## [1] 136
```

ncol(survey_df)

```
## [1] 8
```

3.

4.

5.

6.

7.