SoCalRUG Data Science Hackathon

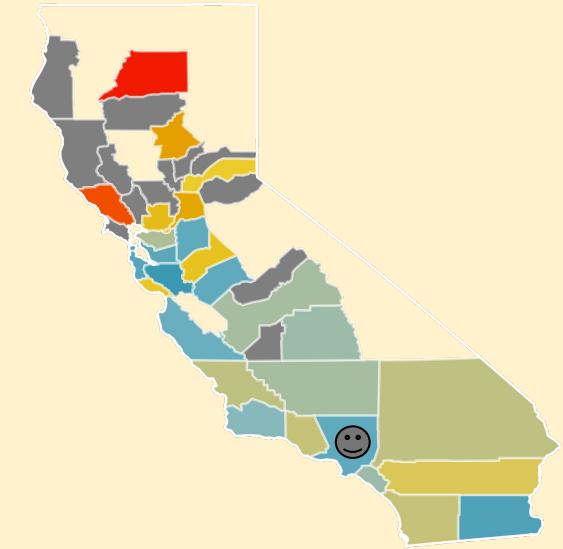
Chris's JETS:

Christopher Chen, Jay Mantuhac, Erica Grabowski, Ted Alexander, Sneha Banerjee

How likely are you to be divorced in California?

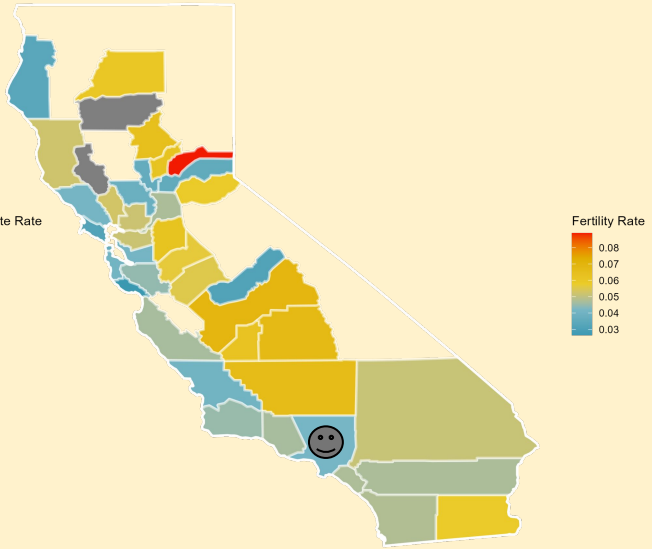
Demographics at the State Level

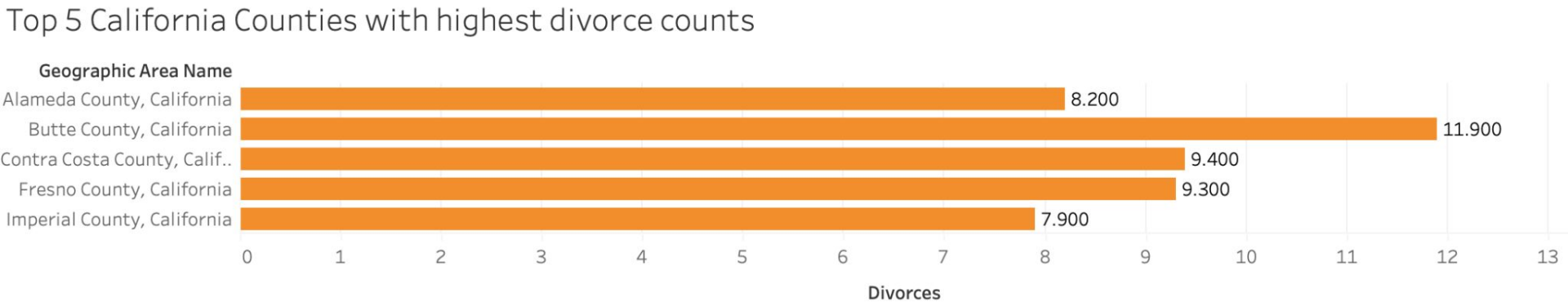
2019 Divorce Rates



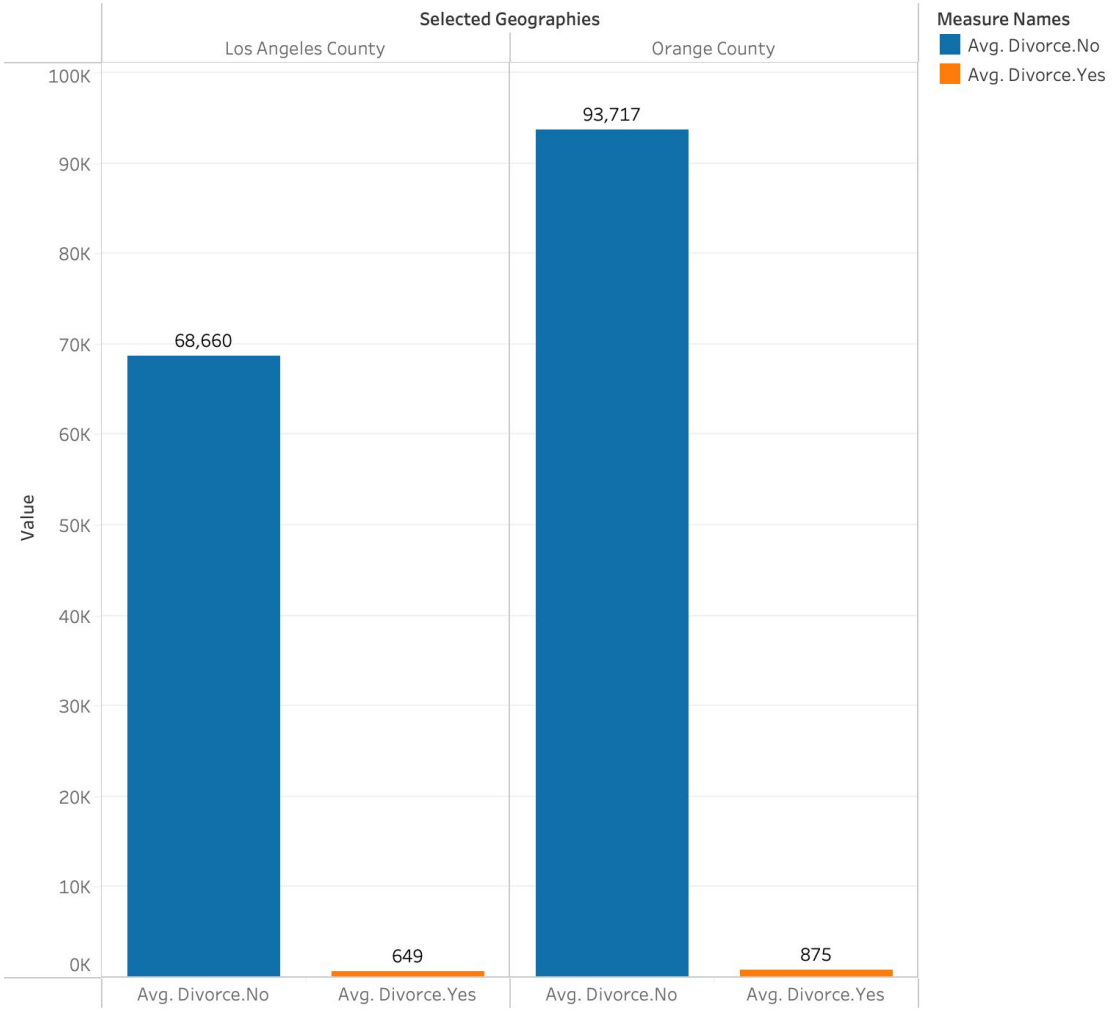
2019 Long Commute Rates

2019 Fertility Rates



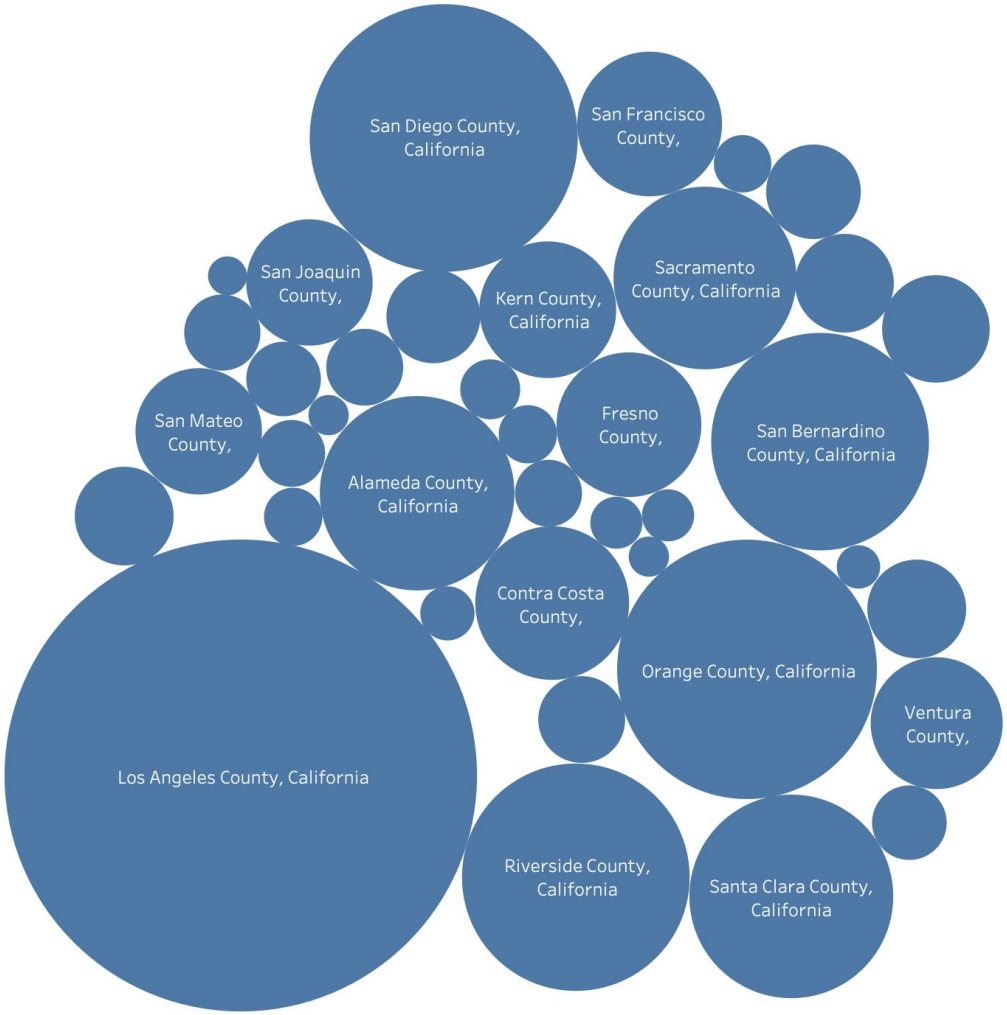

Top 5 California Counties with highest divorce counts
Geographic Area Name
Alameda County, California
Butte County, California
Contra Costa County, Calif..
Fresno County, California
Imperial County, California
8.200
11.900
9.400
9.300
7.900
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Divorces

Average Divorce Counts In LA Vs OC

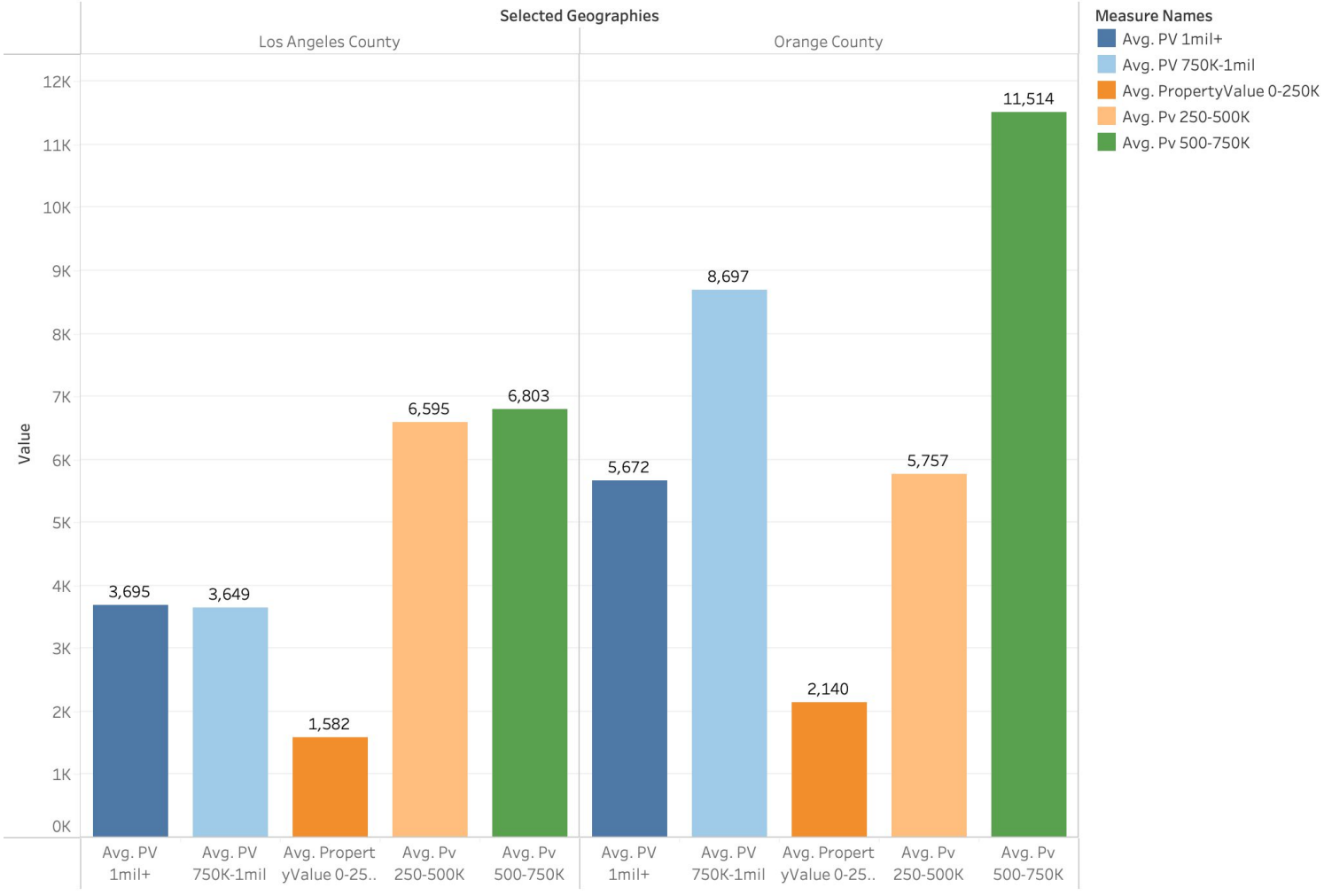


Are you going wrong in these areas in your personal and professional life?

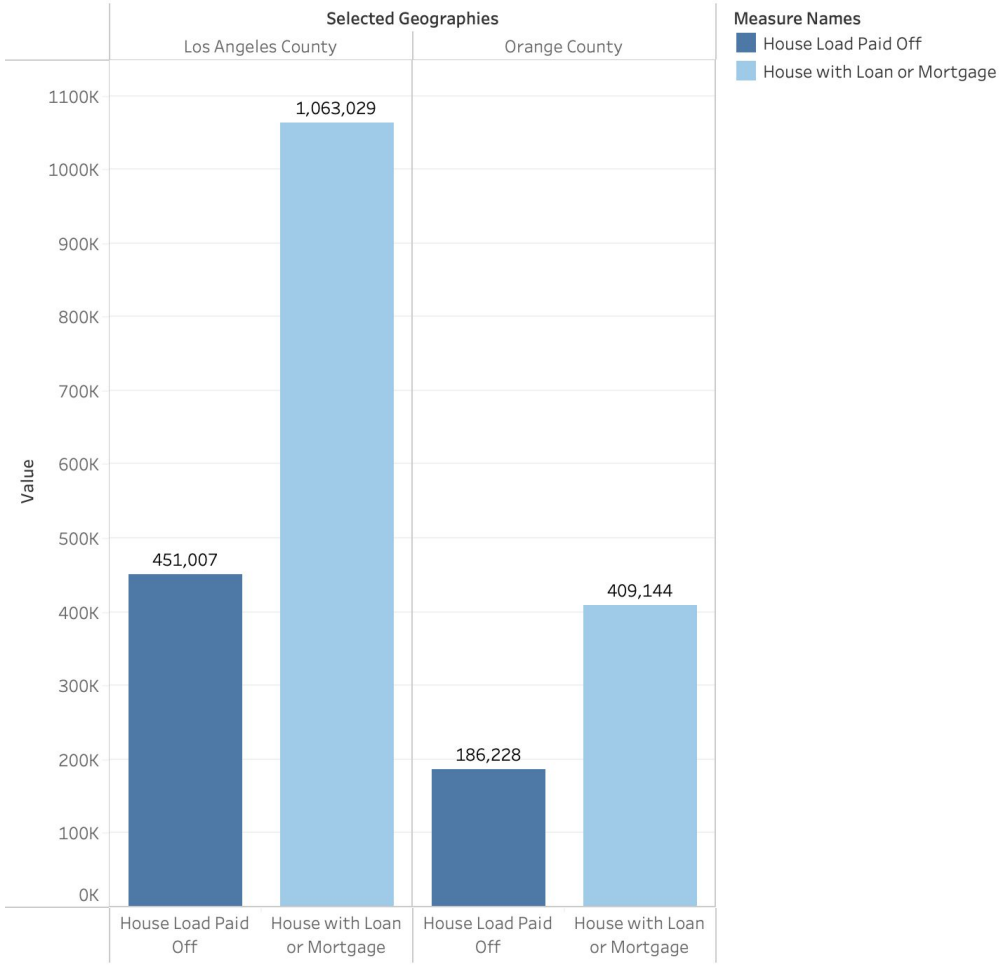
Counties in California with maximum child births



Average number of people in LA vs OC on the basis of property value

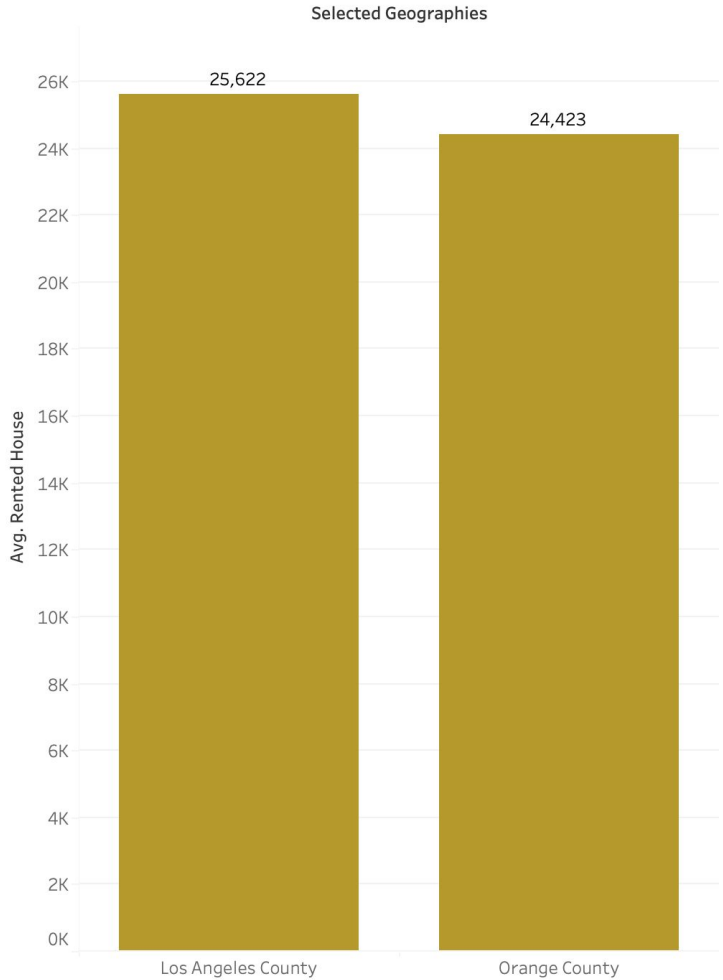


House Owners in LA vs OC



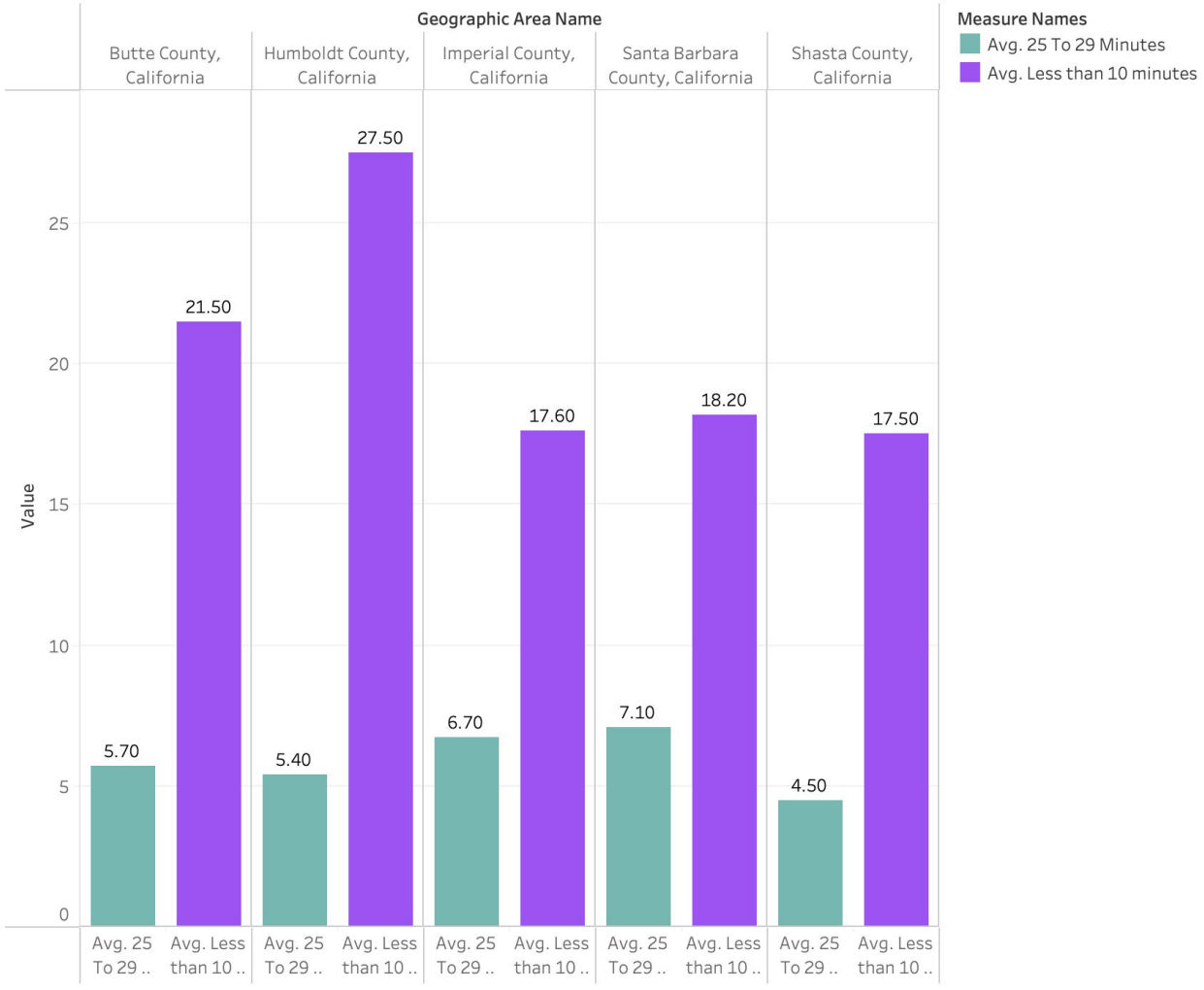
House Load Paid Off and House with Loan or Mortgage for each Selected Geographies.
Color shows details about House Load Paid Off and House with Loan or Mortgage.

Average people renting in LA vs OC

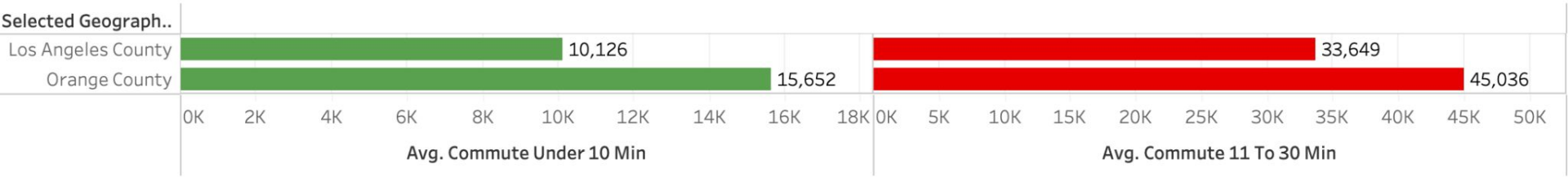


Average of Rented House for each Selected Geographies.

Top 5 California counties based on average time to commute to work

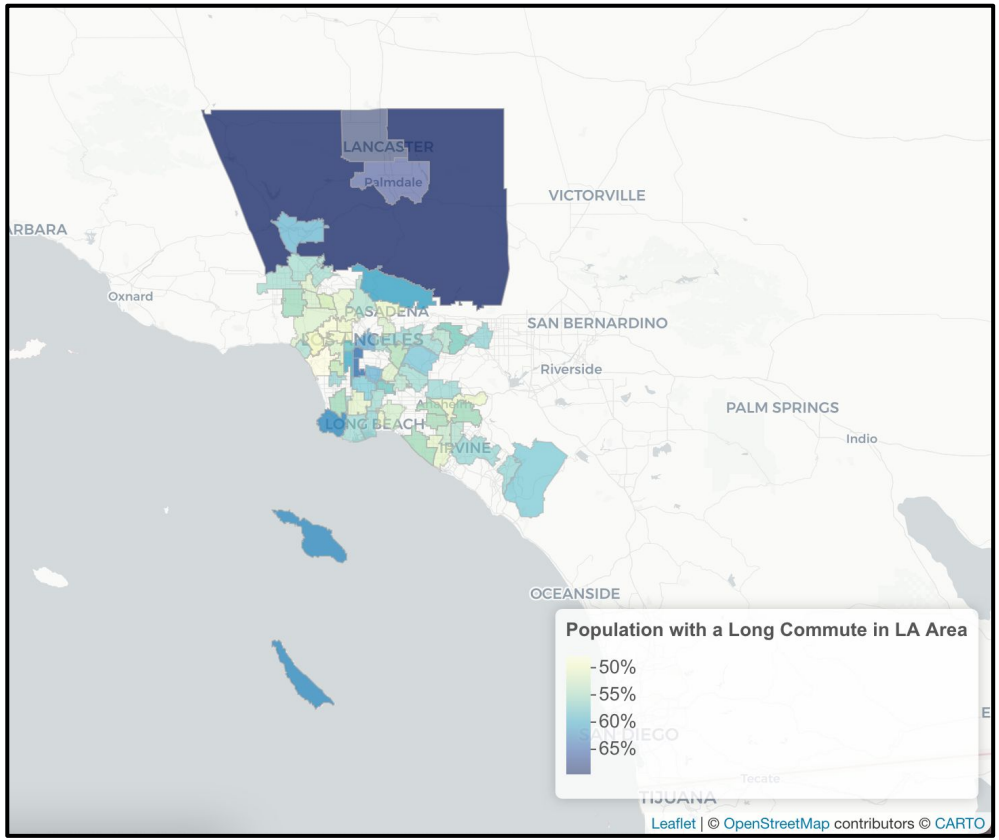


Average number of Commuters in LA vs OC

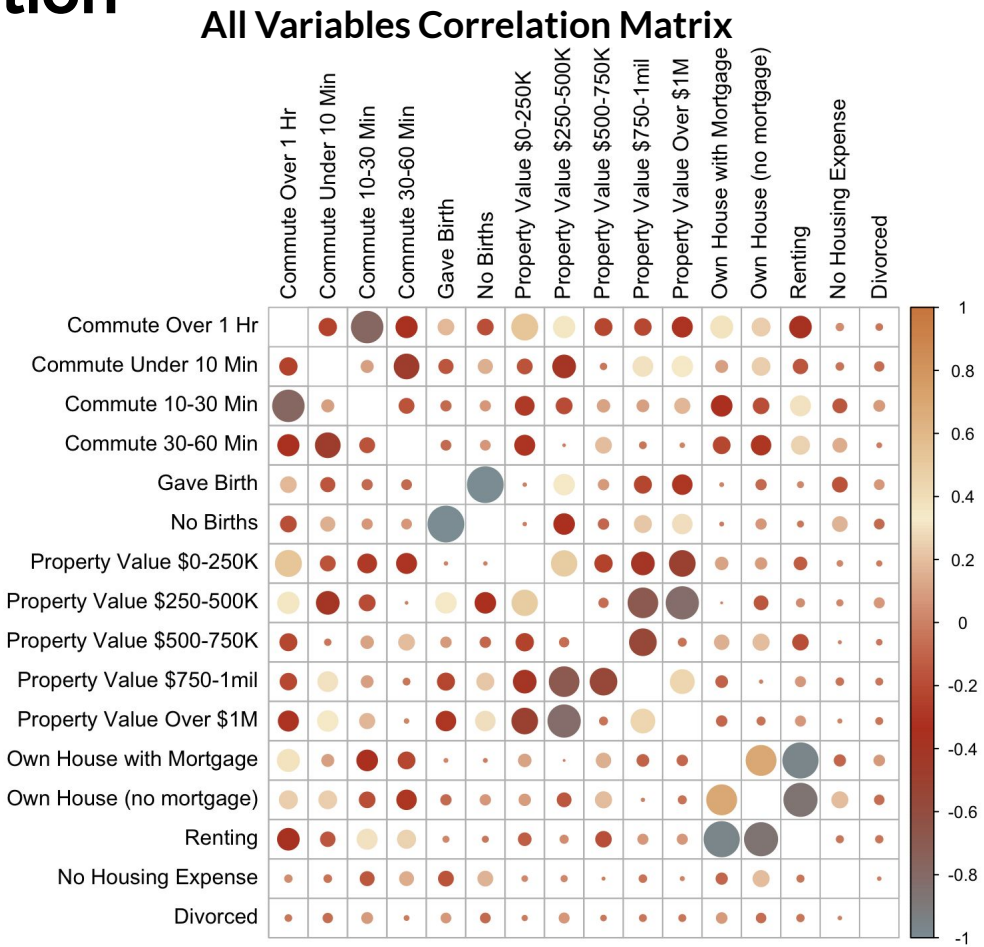


Average of Commute Under 10 Min and average of Commute 11 To 30 Min for each Selected Geographies.

Variation in Commute By Geography (LA Area)



Variable Selection



Higher Divorce Rate Prediction by District Cities (via Cluster Oriented Classification)

Training with K=2 designated cluster(s)

```
DataSet_test_pred <- knn(train = DataSet_train, test = DataSet_test, cl = DataSet_train_labels, k=2)
CrossTable(x = DataSet_test_labels, y = DataSet_test_pred, prop.chisq=FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  22
##
##
##                      | DataSet_test_pred
## DataSet_test_labels |       No  |      Yes  | Row Total |
## --------------------|-----------|-----------|-----------|
##                  No |        16 |         4 |        20 |
##                     |     0.800 |     0.200 |     0.909 |
##                     |     0.889 |     1.000 |           |
##                     |     0.727 |     0.182 |           |
## --------------------|-----------|-----------|-----------|
##                 Yes |         2 |         0 |         2 |
##                     |     1.000 |     0.000 |     0.091 |
##                     |     0.111 |     0.000 |           |
##                     |     0.091 |     0.000 |           |
## --------------------|-----------|-----------|-----------|
##        Column Total |        18 |         4 |        22 |
##                     |     0.818 |     0.182 |           |
## --------------------|-----------|-----------|-----------|
```

Training with K=5 designated cluster(s)

```
DataSet_test_pred <- knn(train = DataSet_train, test = DataSet_test,
CrossTable(x = DataSet_test_labels, y = DataSet_test_pred, prop.chis
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  22
##
##
##                      | DataSet_test_pred
## DataSet_test_labels |       No  | Row Total |
## --------------------|-----------|-----------|
##                  No |        20 |        20 |
##                     |     0.909 |           |
## --------------------|-----------|-----------|
##                 Yes |         2 |         2 |
##                     |     0.091 |           |
## --------------------|-----------|-----------|
##        Column Total |        22 |        22 |
## --------------------|-----------|-----------|
```

Thank You