

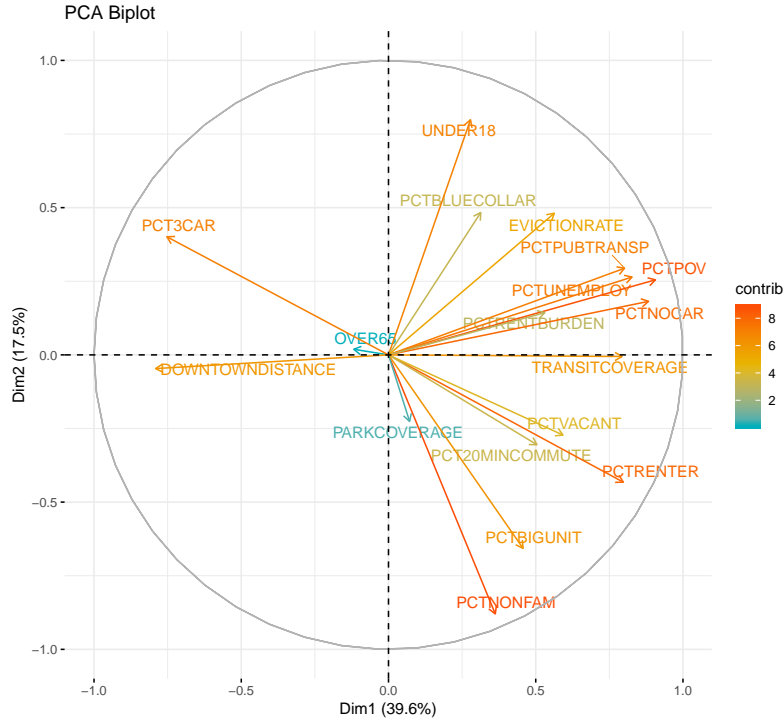
Logistic Regression Figures

David Knorr

March 2019

Logistic Regression

```
## Joining, by = "CensusTractID"
```



```
## \begin{group}\fontsize{11}{13}\selectfont
##
## \begin{longtable}{lrr}
## \caption{\label{tab:dk3}Confusion Matrix}\\
## \toprule
## \multicolumn{1}{c}{Predicted} & \multicolumn{2}{c}{Observed} \\
## \cmidrule{1}{3pt}{1-1} \cmidrule{1}{3pt}{2-3}
## & Not Gentrifying & Gentrifying\\
## \midrule
## \endfirsthead
## \caption[]{}{Confusion Matrix \textit{(continued)}}\\
## \toprule
## & Not Gentrifying & Gentrifying\\
## \midrule
## \endhead
## \
## \endfoot
## \bottomrule
## \endlastfoot
## Not Gentrifying & 31 & 1\\
## Gentrifying & 8 & 5\\
## \end{longtable}
## \end{group}

## \begin{group}\fontsize{11}{13}\selectfont
##
## \begin{longtable}{lr}
## \caption{\label{tab:dk3}Model Performance Metrics}\\
## \toprule
## & value\\
## \midrule
```

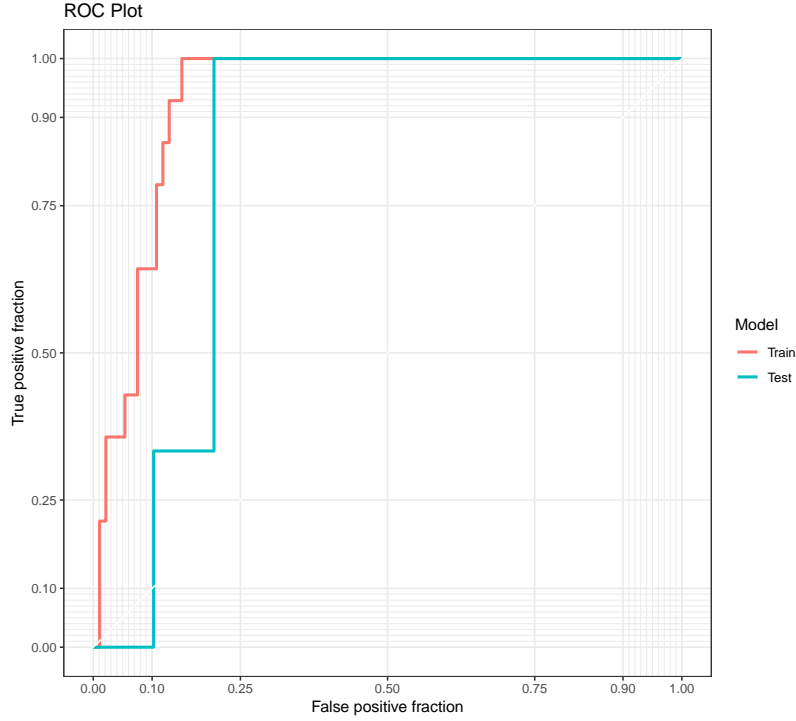
```

## \endfirsthead
## \caption[]{{Model Performance Metrics \textit{(continued)}}}\
## \toprule
##   & value\\
## \midrule
## \endhead
## \
## \endfoot
## \bottomrule
## \endlastfoot
## Sensitivity & 0.795\\
## Specificity & 0.833\\
## Pos Pred Value & 0.969\\
## Neg Pred Value & 0.385\\
## Precision & 0.969\\
## Recall & 0.795\\
## F1 & 0.873\\
## Prevalence & 0.867\\
## Detection Rate & 0.689\\
## Detection Prevalence & 0.711\\
## Balanced Accuracy & 0.814\\*
## \end{longtable}
## \endgroup{}

```

Table 1: PCA Rotation Values

Variables	PC1	PC2	PC3	PC4	PC5
PCTRENTER	0.299	-0.243	0.247	-0.009	0.048
PCTVACANT	0.222	-0.153	-0.007	-0.301	-0.166
PCTNONFAM	0.136	-0.496	0.004	-0.137	-0.076
UNDER18	0.104	0.45	0.185	0.268	0.148
OVER65	-0.044	0.011	-0.578	-0.086	0.187
PCTPOV	0.34	0.144	0.02	0.132	0.102
PCTUNEMPLOY	0.31	0.149	0.024	0.11	-0.063
PCTNOCAR	0.331	0.102	-0.011	0.225	0.154
PCT3CAR	-0.282	0.226	-0.159	0.021	-0.153
PCTPUBTRANSP	0.3	0.166	-0.01	0.22	0.062
PCT20MINCOMMUTE	0.189	-0.172	-0.412	0.278	-0.288
PCTBLUECOLLAR	0.117	0.272	0.127	-0.477	-0.084
PCTBIGUNIT	0.171	-0.37	0.332	0.095	0.121
PCTRENTBURDEN	0.199	0.081	-0.221	-0.462	0.089
PARKCOVERAGE	0.027	-0.127	-0.197	-0.08	0.82
DOWNTOWNDISTANCE	-0.296	-0.026	0.361	-0.094	0.157
TRANSITCOVERAGE	0.297	-0.003	-0.118	-0.092	-0.18
EVICTONRATE	0.211	0.271	0.13	-0.364	0.006



Area under the curve: 0.8291

Area under the curve: 0.9309

Table 2: Random Forest Model Performance

Metric	Value
Accuracy	0.800
Kappa	0.421
AccuracyLower	0.654
AccuracyUpper	0.904
Sensitivity	0.795
Specificity	0.833
Precision	0.969
Recall	0.795
F1	0.873
Balanced Accuracy	0.814

Importance of components:

##	PC1	PC2	PC3	PC4	PC5	PC6
## Standard deviation	2.6693	1.7751	1.34024	1.07548	1.03246	0.84920
## Proportion of Variance	0.3958	0.1751	0.09979	0.06426	0.05922	0.04006
## Cumulative Proportion	0.3958	0.5709	0.67068	0.73494	0.79416	0.83422
##	PC7	PC8	PC9	PC10	PC11	PC12
## Standard deviation	0.8095	0.75679	0.66454	0.60043	0.5002	0.45665
## Proportion of Variance	0.0364	0.03182	0.02453	0.02003	0.0139	0.01159
## Cumulative Proportion	0.8706	0.90244	0.92697	0.94700	0.9609	0.97249
##	PC13	PC14	PC15	PC16	PC17	PC18
## Standard deviation	0.43133	0.3674	0.2547	0.21293	0.19063	0.16613

```
## Proportion of Variance 0.01034 0.0075 0.0036 0.00252 0.00202 0.00153
## Cumulative Proportion 0.98283 0.9903 0.9939 0.99645 0.99847 1.00000
```

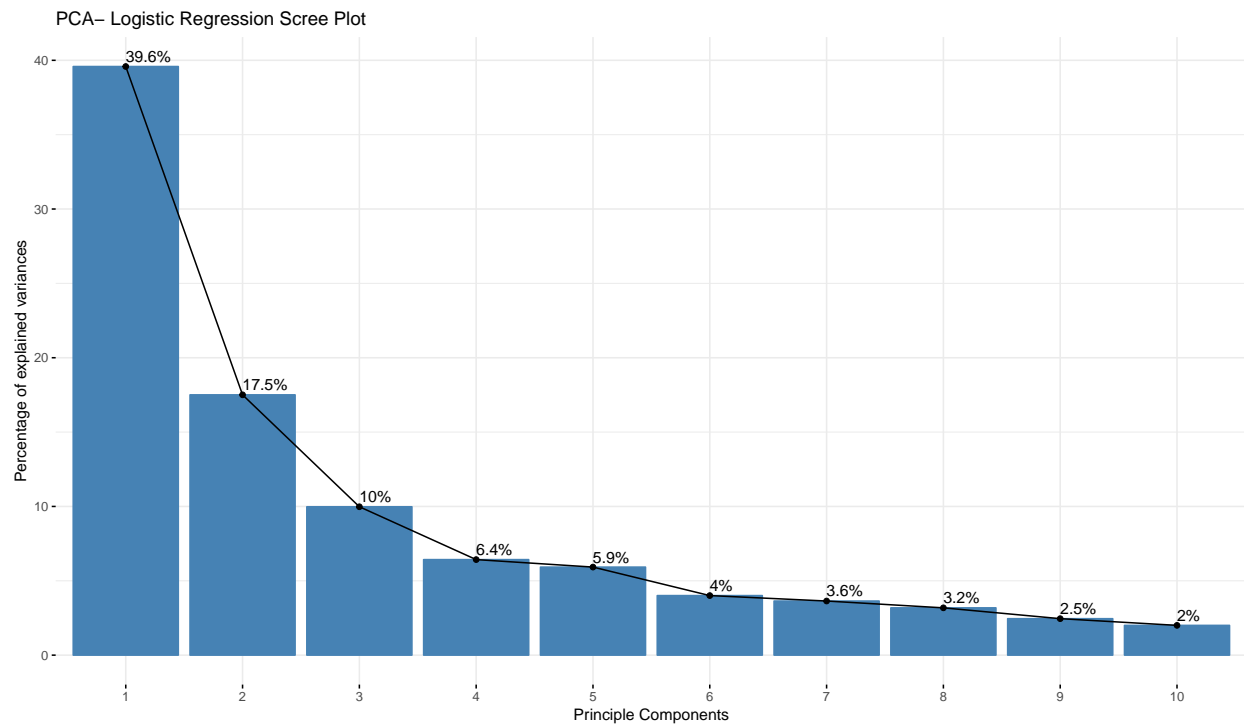


Table 3: Logistic Classification Model Parameters

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.501	0.79	-4.41	0.0000
PC1	0.716	0.18	3.88	0.0001
PC2	-0.152	0.20	-0.77	0.4436
PC3	-1.057	0.45	-2.35	0.0189
PC4	0.150	0.33	0.45	0.6523
PC5	-0.241	0.38	-0.63	0.5284