Thesis Work: Part 3

Emma Lasky

Overview of Categorization

We categorized the slope of each zip code into:

- Significant and negative (0)
- Not significant (1)
- Significant and positive (2)

We aimed to identify slopes that were significant and negative and ones that were not significant to illustrate zip codes that experienced no LST change over time (1) versus those that had a decrease in LST over time (0)

Number of Zip Codes per City

Sacramento = 29

Oakland = 14

San Francisco = 28

San Jose = 28

Fresno = 18

Los Angeles = 64

San Diego = 35

Method

I chose to visualize four different p-value thresholds:

p-val = 0.05, 0.10, 0.15, & 0.25

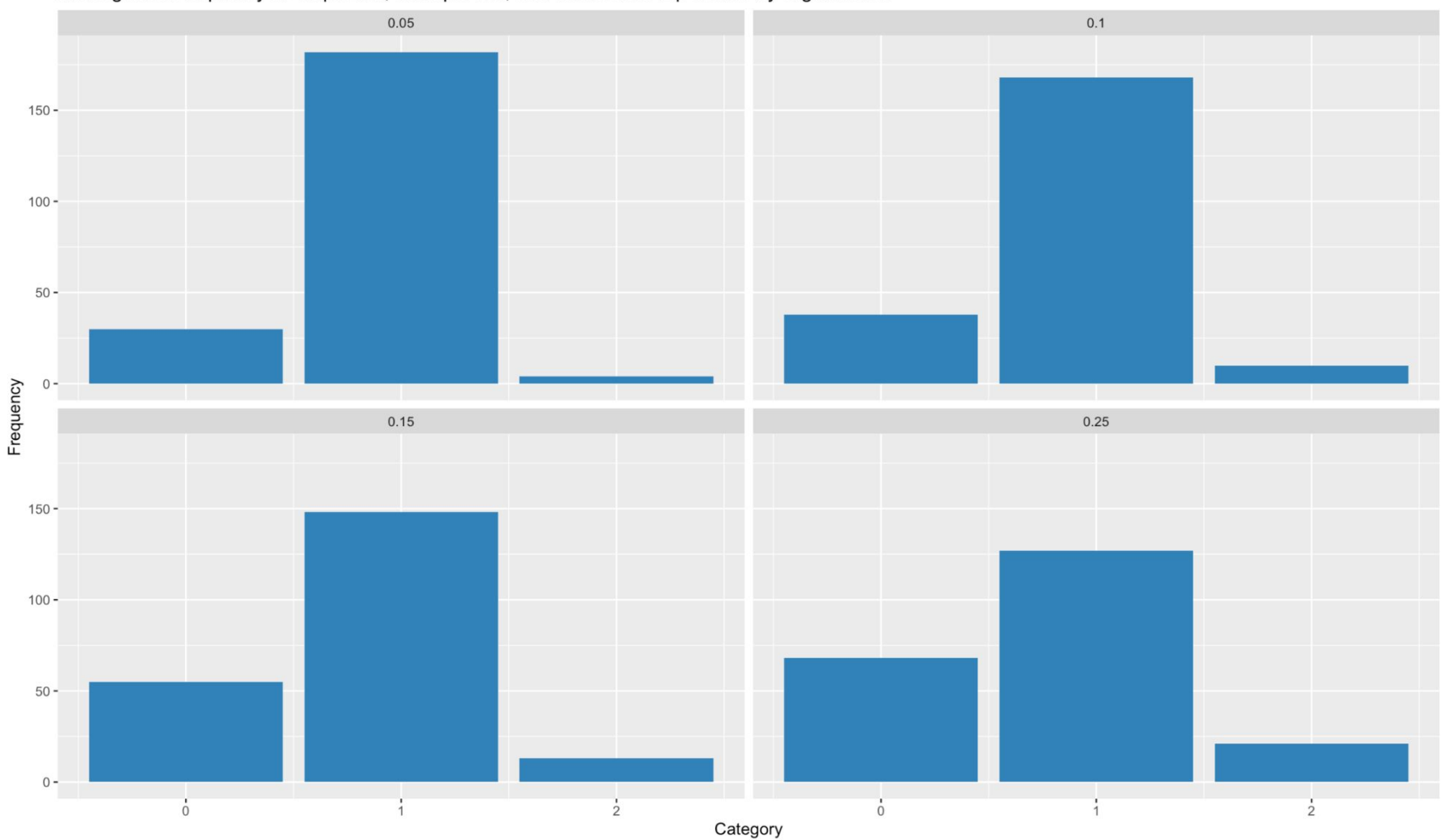
By adjusting the threshold, I was able to identify more conservative categorizations of improved, unimproved, and worsened zip codes, which increases the number of 0, 1 zip codes that can be matched

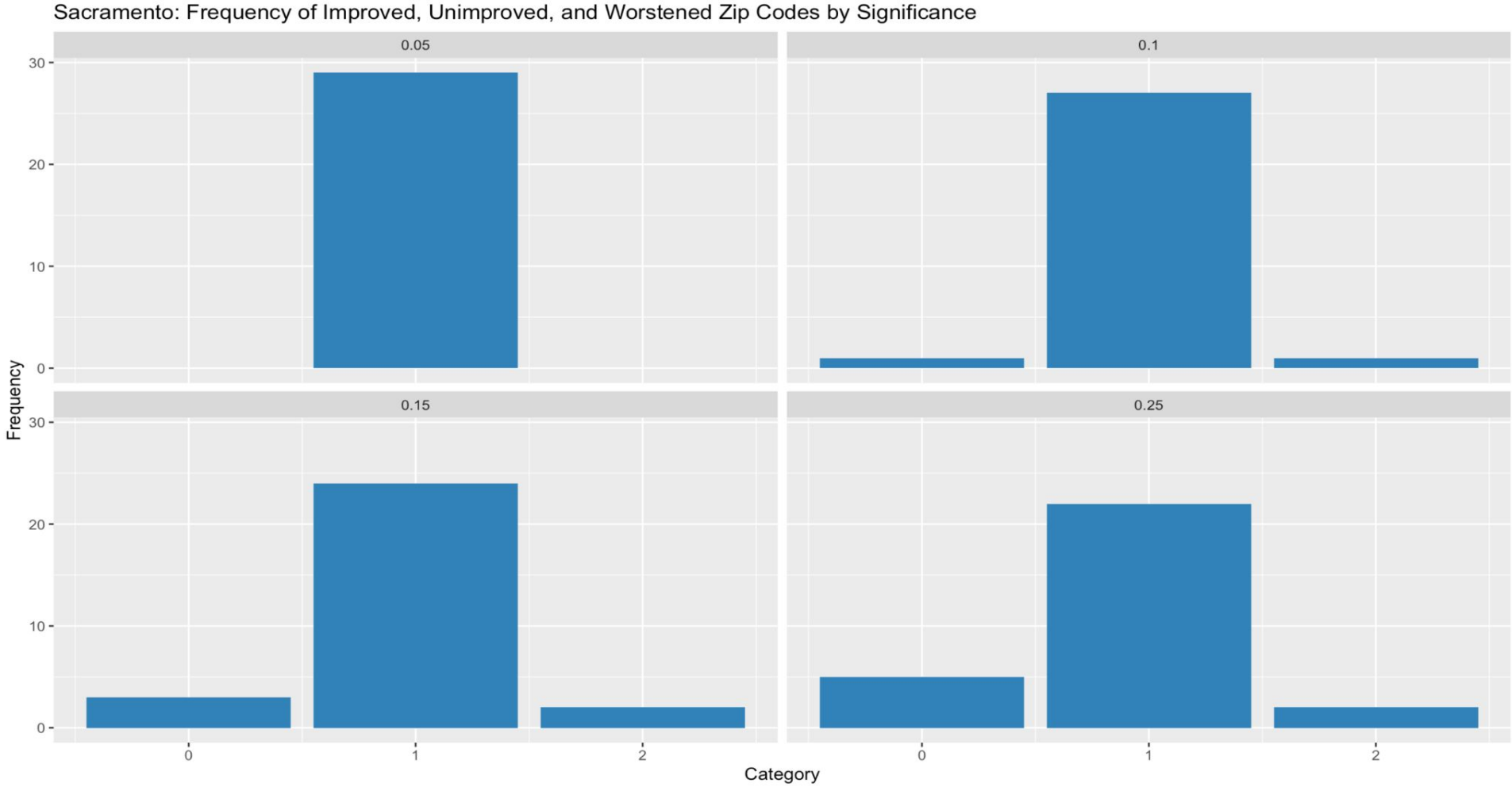
Method

To categorize each zip code, I relied on the case_when() and if_else() *plyr* and *dplyr* functions

```
lm_reg3oak$pos = ifelse(lm_reg3oak$estimate>=0.0000,1,0)
lm_reg3sf$pos = ifelse(lm_reg3sf$estimate>=0.0000,1,0)
lm_reg3sj$pos = ifelse(lm_reg3sj$estimate>=0.0000,1,0)
```

```
cat_reg3oak <- lm_reg3oak %>% mutate(status05 = case_when(
  p.value <= 0.05 & pos == 1 ~ 2,
  p.value <= 0.05 & pos == 0 ~ 0,
  p.value > 0.05 ~ 1
)) %>%
```


All Regions: Frequency of Improved, Unimproved, and Worstened Zip Codes by Significance
0.05
0.1
0.15
0.25
Frequency
0
50
100
150
Category
0
1
2

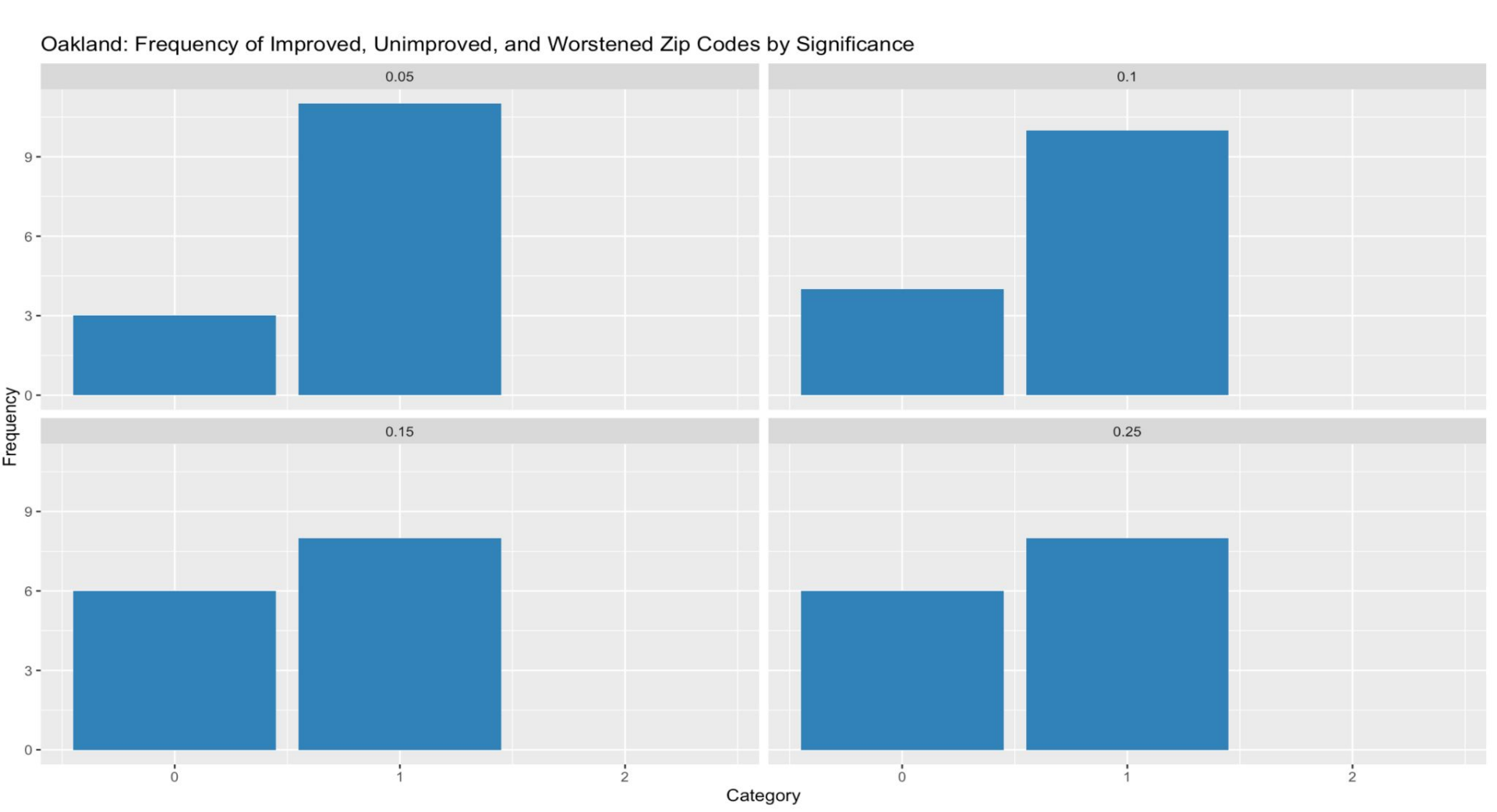

Sacramento: Frequency of Improved, Unimproved, and Worstened Zip Codes by Significance
0.05
0.1
0.15
0.25
Frequency
30
20
10
0
0
1
2
Category

	ZCTA	estimate	std.error	statistic	p.value	pos	status25	status15	status10	status05
1	95811	-0.015041800	0.026565017	-0.56622589	0.58892426	0	1	1	1	1
2	95814	-0.013831230	0.024086456	-0.57423266	0.58378311	0	1	1	1	1
3	95815	-0.013077700	0.029170977	-0.44831204	0.66747559	0	1	1	1	1
4	95816	0.021657460	0.029572337	0.73235539	0.48774252	1	1	1	1	1
5	95817	-0.001409617	0.038317217	-0.03678809	0.97155538	0	1	1	1	1
6	95818	-0.003504324	0.030401503	-0.11526813	0.91107334	0	1	1	1	1
7	95819	-0.001858788	0.035251669	-0.05272907	0.95924069	0	1	1	1	1
8	95820	0.010944193	0.035731799	0.30628721	0.76720557	1	1	1	1	1

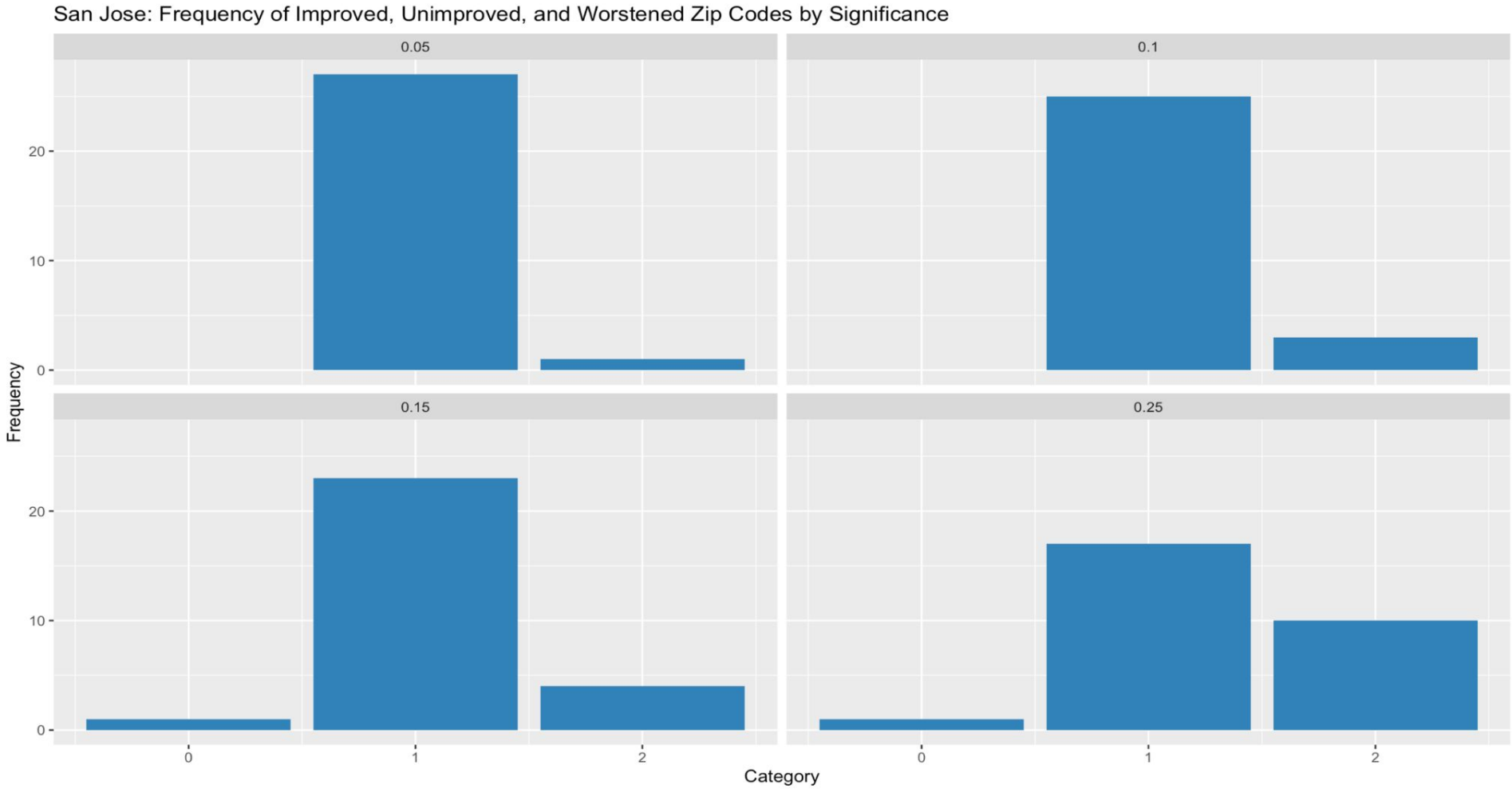
estimate = slope of LST

pos = trajectory of slope (positive 1, negative 0)

status25, 15, 10, 05 = p.value above or below designated p-value for that column


Oakland: Frequency of Improved, Unimproved, and Worstened Zip Codes by Significance
0.05
0.1
0.15
0.25
Frequency
0
3
6
9
0
1
2
Category

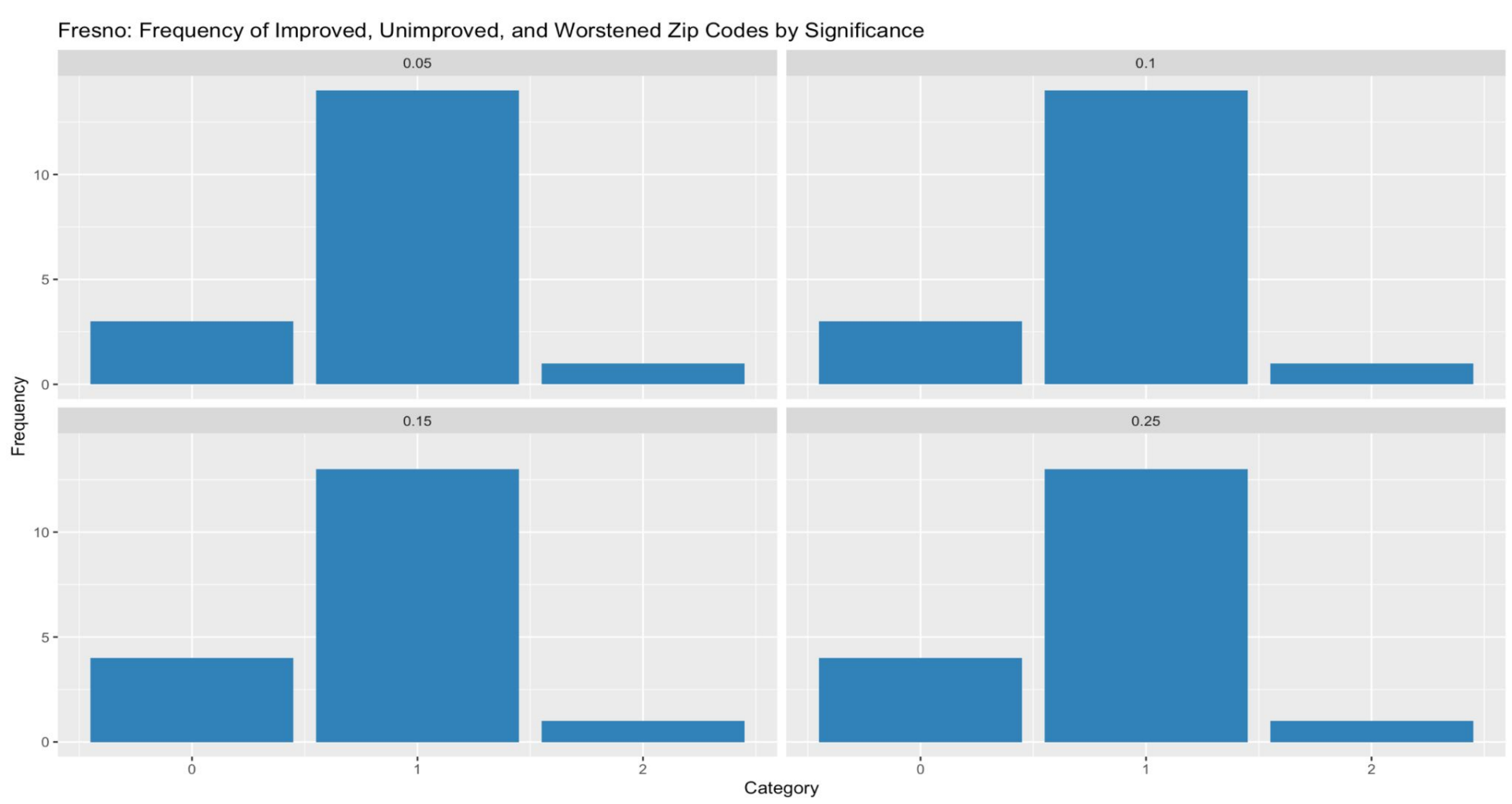
	ZCTA	estimate	std.error	statistic	p.value	pos	status05	status10	status15	status25
1	94601	-0.025744071	0.02398673	-1.07326320	0.318752259	0	1	1	1	1
2	94602	-0.017536006	0.02131566	-0.82268181	0.442149850	0	1	1	1	1
3	94603	-0.031107844	0.04479708	-0.69441671	0.509804109	0	1	1	1	1
4	94605	-0.013648914	0.01457757	-0.93629581	0.380289643	0	1	1	1	1
5	94606	-0.023018889	0.04009651	-0.57408703	0.586769718	0	1	1	1	1
6	94607	-0.054090985	0.01719381	-3.14595676	0.016243490	0	0	0	0	0
7	94609	-0.043390417	0.02184732	-1.98607465	0.087395610	0	1	0	0	0


San Jose: Frequency of Improved, Unimproved, and Worstened Zip Codes by Significance
0.05
0.1
0.15
0.25
Frequency
0
10
20
0
1
2
Category

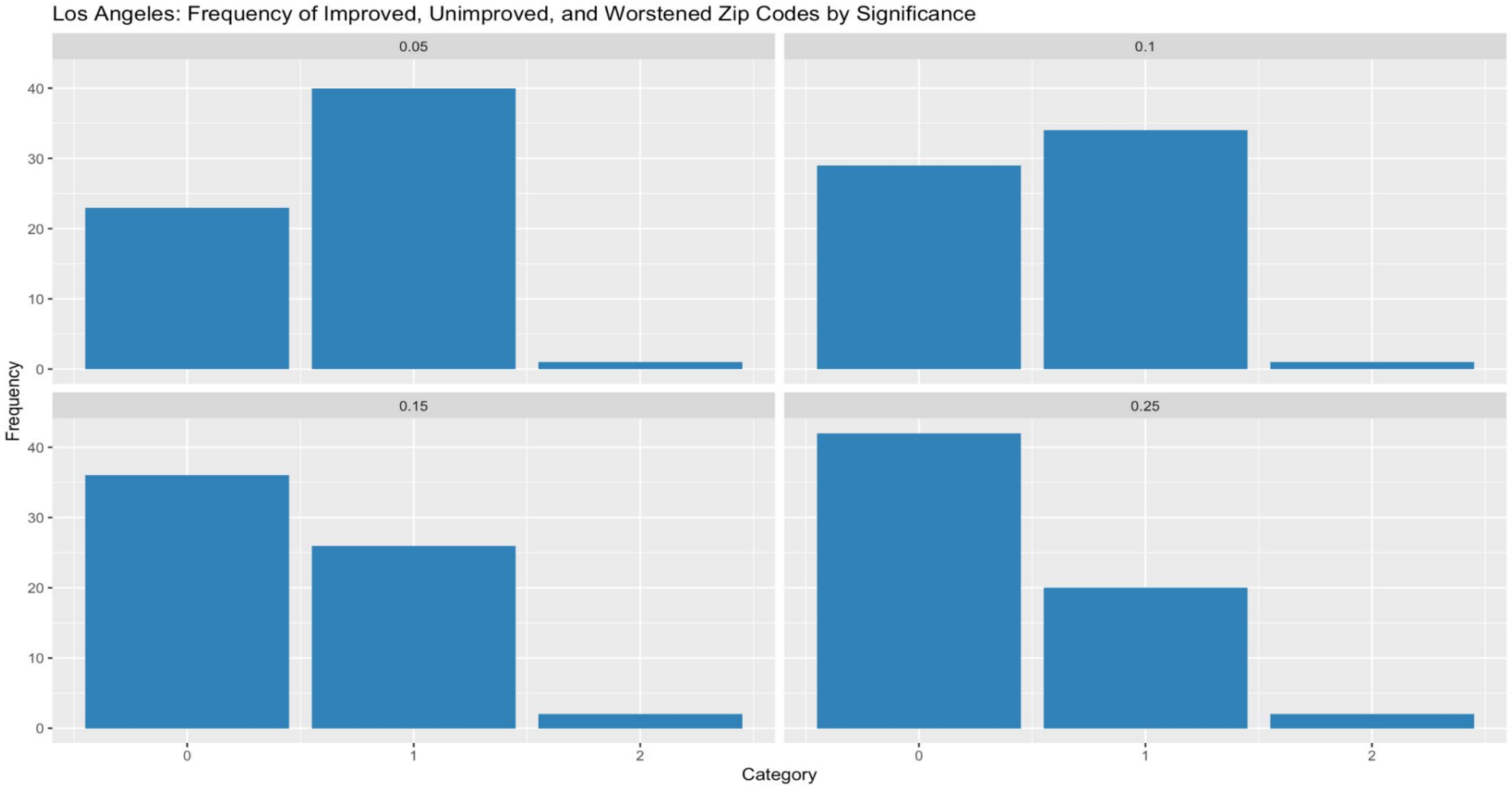
	ZCTA	estimate	std.error	statistic	p.value	pos	status05	status10	status15	status25
1	95110	0.006328237	0.023649421	0.26758529	0.79673142	1	1	1	1	1
2	95111	0.019309902	0.017712101	1.09020955	0.30736629	1	1	1	1	1
3	95112	0.017590266	0.033444190	0.52595880	0.61318090	1	1	1	1	1
4	95113	0.006328237	0.023649421	0.26758529	0.79673142	1	1	1	1	1
5	95116	0.015606119	0.025692900	0.60740980	0.55859081	1	1	1	1	1
6	95117	0.002897339	0.036484023	0.07941391	0.93865397	1	1	1	1	1
7	95118	0.015148588	0.021473537	0.70545376	0.50054970	1	1	1	1	1
8	95119	-0.028920269	0.016279072	-1.77653051	0.11354984	0	1	1	0	0
9	95120	-0.015923379	0.019971913	-0.79728860	0.44829817	0	1	1	1	1


San Francisco: Frequency of Improved, Unimproved, and Worstened Zip Codes by Significance
0.05
0.1
0.15
0.25
Frequency
0
10
20
0
1
2
Category

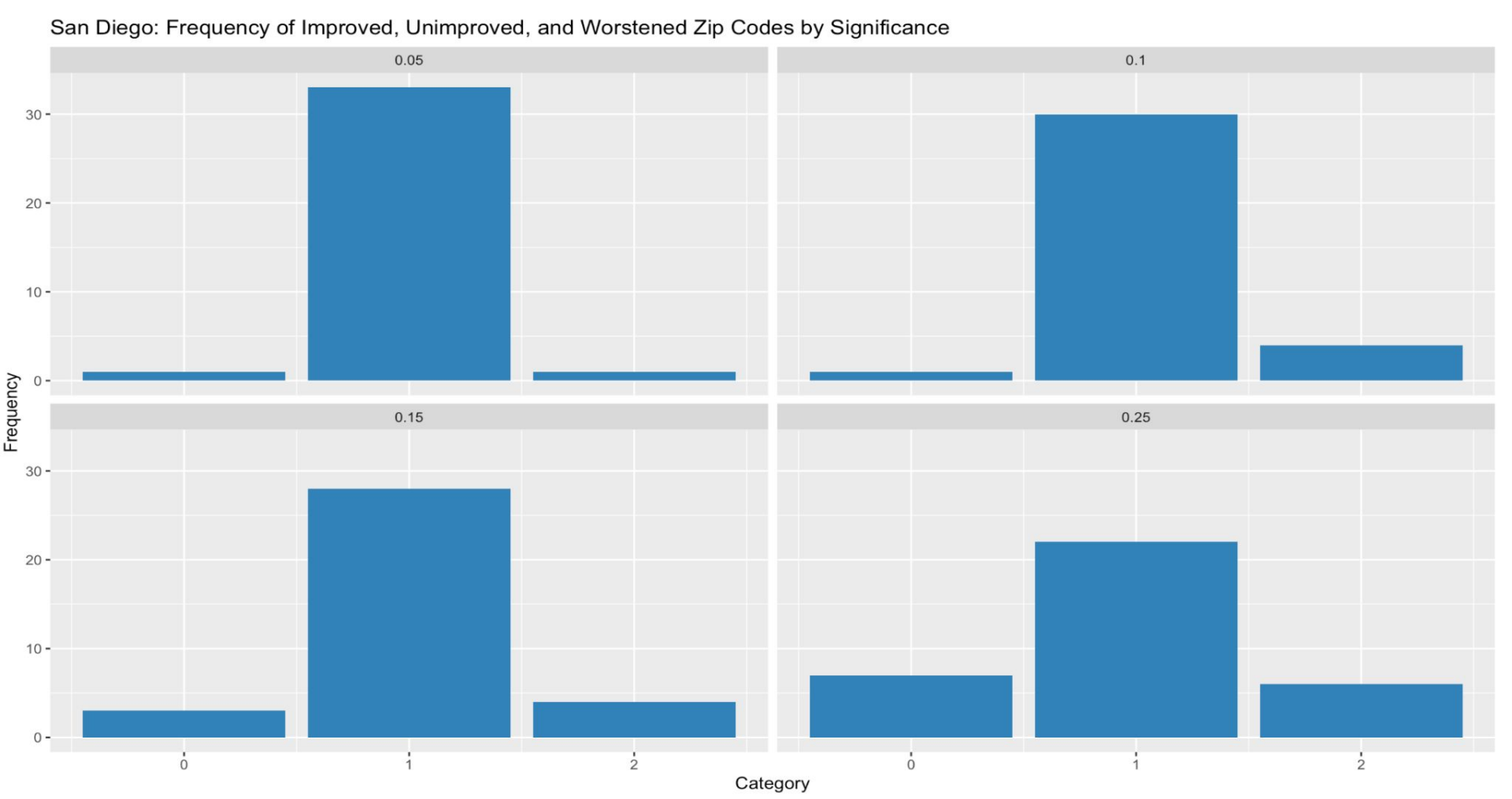
	ZCTA	estimate	std.error	statistic	p.value	pos	status05	status10	status15	status25
1	94102	0.006273840	0.015793606	0.3972392	0.7049368	1	1	1	1	1
2	94103	0.002611353	0.021589456	0.1209550	0.9071254	1	1	1	1	1
3	94104	-0.003243952	0.019417220	-0.1670657	0.8728066	0	1	1	1	1
4	94105	-0.008709423	0.014618380	-0.5957858	0.5730906	0	1	1	1	1
5	94107	-0.009050844	0.028694479	-0.3154211	0.7631245	0	1	1	1	1
6	94108	-0.012409705	0.020324167	-0.6105886	0.5638696	0	1	1	1	1
7	94109	-0.013602743	0.022056317	-0.6167277	0.5600720	0	1	1	1	1
8	94110	-0.003200211	0.025627241	-0.1248754	0.9047010	0	1	1	1	1
9	94111	-0.005908158	0.021004578	-0.2812795	0.7879446	0	1	1	1	1
10	94112	-0.022819578	0.012458412	-1.8316602	0.1043712	0	1	1	0	0


Fresno: Frequency of Improved, Unimproved, and Worstened Zip Codes by Significance
0.05
0.1
0.15
0.25
Frequency
0
5
10
0
1
2
Category

	ZCTA	estimate	std.error	statistic	p.value	pos	status05	status10	status15	status25
1	93650	-0.0320705862	0.007704467	-4.16259637	0.002438287	0	0	0	0	0
2	93701	-0.0010700119	0.006256707	-0.17101838	0.867992459	0	1	1	1	1
3	93702	-0.0055127145	0.008237341	-0.66923467	0.520142220	0	1	1	1	1
4	93703	-0.0008593308	0.009229739	-0.09310456	0.927860152	0	1	1	1	1
5	93704	0.0021850494	0.006432297	0.33969971	0.741875693	1	1	1	1	1
6	93705	0.0019961518	0.006148799	0.32464093	0.752871503	1	1	1	1	1
7	93706	-0.0199091429	0.006739471	-2.95411064	0.016112945	0	0	0	0	0
8	93710	0.0045420242	0.008401285	0.54063449	0.603481874	1	1	1	1	1


Los Angeles: Frequency of Improved, Unimproved, and Worstened Zip Codes by Significance
0.05
0.1
0.15
0.25
Frequency
0
10
20
30
40
0
1
2
Category

	ZCTA	estimate	std.error	statistic	p.value	pos	status05	status10	status15	status25
1	90001	-0.0258395787	0.016190391	-1.59598242	1.491594e-01	0	1	1	0	0
2	90002	-0.0246322226	0.018586210	-1.32529563	2.216706e-01	0	1	1	1	0
3	90003	-0.0184804987	0.025778506	-0.71689564	4.966556e-01	0	1	1	1	1
4	90004	-0.0273973908	0.014702635	-1.86343410	9.940487e-02	0	1	0	0	0
5	90005	-0.0275589437	0.012962149	-2.12610912	6.619272e-02	0	1	0	0	0
6	90006	-0.0316147752	0.011383904	-2.77714711	2.403012e-02	0	0	0	0	0
7	90007	-0.0298711097	0.012250804	-2.43829798	4.067234e-02	0	0	0	0	0
8	90008	-0.0543256954	0.021211463	-2.56114802	3.358692e-02	0	0	0	0	0
9	90010	-0.0275589437	0.012962149	-2.12610912	6.619272e-02	0	1	0	0	0


San Diego: Frequency of Improved, Unimproved, and Worstened Zip Codes by Significance
0.05
0.1
0.15
0.25
Frequency
0
10
20
30
0
1
2
Category

	ZCTA	estimate	std.error	statistic	p.value	pos	status05	status10	status15	status25
1	92101	-0.062876640	0.02427209	-2.59049110	0.029189951	0	0	0	0	0
2	92102	-0.026513440	0.02015180	-1.31568606	0.224730823	0	1	1	1	0
3	92103	0.011247554	0.02826145	0.39798218	0.702501702	1	1	1	1	1
4	92104	0.021152912	0.02619679	0.80746204	0.442744212	1	1	1	1	1
5	92105	0.010061475	0.03057607	0.32906373	0.751733831	1	1	1	1	1
6	92106	-0.040815802	0.02943807	-1.38649714	0.203008302	0	1	1	1	0
7	92107	-0.057870755	0.03278692	-1.76505604	0.115553453	0	1	1	0	0
8	92108	0.004639029	0.03145622	0.14747571	0.886914457	1	1	1	1	1
9	92109	0.043219058	0.02047272	2.11105571	0.079257076	1	1	2	2	2

Next Steps

1. Choose a significance threshold
2. Decide on ACS covariate data we will match on
    a. I have downloaded ACS 5 year economic, social, housing, and demographic and housing estimates from 2017
    b. What specific covariates do we want to match on?
        i. ACS data can be quite specific. For example, I imagine we want to match on economic status, but not the percent of households with a child in nursery school
        ii. Because of how few zip codes there are in each city, can I refine the covariates to ~15?
            1. Maybe like these?
        iii. What is the calibration we will use when looking at probabilities of each zip code?
            1. Will we drop unmatched zip codes?