# Data Memo for Al

## Summary statistics

### Import taxlots

### Clean

```
df %<>%
  mutate(proud_flag =  grepl("PROUD", OWNER1) | grepl("PROUD", OWNER2) | grepl("PROUD", OWNER3),
         trust_flag = grepl("TRUST", OWNER1) | grepl("TRUST", OWNER2) | grepl("TRUST", OWNER3),
         top_1 =  SALEPRICE > quantile(SALEPRICE, .99),
         price_diff = SALEPRICE - LANDVAL3,
         price_ratio = SALEPRICE/LANDVAL3 * 100,
         vacant_dummy = PRPCD_DESC == "VACANT LAND") %>%
  mutate(arms_length = price_ratio > 20)

constraints <- c("conWetland", "conNatAm",
                   "conAirHgt", "conCovrly", "conPovrly", "conHeliprt",
                   "conHist", "conHistLdm", "conInstit", "conLSHA", "conLUST",
                   "conNoise", "conPrvCom", "conSewer", "conSLIDO",
                   "conSlp25", "conStorm", "conTranCap", "conTranSub",
                   "conTranInt", "conTranSub", "conView", "conWater",
                 "conGW", "conPubOwn", "conFldway", "conFld100", "conECSI")



# switch the NAs in the constraints to 0s
to0 <- function(x){ifelse(is.na(x), 0, x)}


trim <- df %>%
  filter(proud_flag == F & top_1 == F &
           arms_length == T & vacant_dummy == F) %>%
  mutate_at(vars(constraints), to0)

constraint_sums <- trim %>%
  select(constraints) %>%
  rowSums()

trim %<>%
  mutate(is_constrained = constraint_sums > 0)
```

### Tables

Table 1: Constraint frequency by property type

| | Mixed Use (N=1657) | Multi-family (N=4385) | Non-conforming (N=437) | Single-family (N=25752) | Total (N=32231) | p value |
|---|---|---|---|---|---|---|
| **conWetland** | | | | | | < 0.001 |
| 0 | 1657 (100.0%) | 4382 (99.9%) | 429 (98.2%) | 25730 (99.9%) | 32198 (99.9%) | |
| 1 | 0 (0.0%) | 3 (0.1%) | 8 (1.8%) | 22 (0.1%) | 33 (0.1%) | |
| **conNatAm** | | | | | | < 0.001 |
| 0 | 1657 (100.0%) | 4385 (100.0%) | 437 (100.0%) | 25752 (100.0%) | 32231 (100.0%) | |
| **conAirHgt** | | | | | | < 0.001 |
| 0 | 1425 (86.0%) | 3726 (85.0%) | 285 (65.2%) | 21120 (82.0%) | 26556 (82.4%) | |
| 1 | 232 (14.0%) | 659 (15.0%) | 152 (34.8%) | 4632 (18.0%) | 5675 (17.6%) | |
| **conCovrly** | | | | | | < 0.001 |
| 0 | 1633 (98.6%) | 4326 (98.7%) | 400 (91.5%) | 24587 (95.5%) | 30946 (96.0%) | |
| 1 | 24 (1.4%) | 59 (1.3%) | 37 (8.5%) | 1165 (4.5%) | 1285 (4.0%) | |
| **conPovrly** | | | | | | < 0.001 |
| 0 | 1657 (100.0%) | 4381 (99.9%) | 425 (97.3%) | 25274 (98.1%) | 31737 (98.5%) | |
| 1 | 0 (0.0%) | 4 (0.1%) | 12 (2.7%) | 478 (1.9%) | 494 (1.5%) | |
| **conHeliprt** | | | | | | < 0.001 |
| 0 | 1657 (100.0%) | 4385 (100.0%) | 437 (100.0%) | 25752 (100.0%) | 32231 (100.0%) | |
| **conHist** | | | | | | < 0.001 |
| 0 | 1563 (94.3%) | 4138 (94.4%) | 435 (99.5%) | 25081 (97.4%) | 31217 (96.9%) | |
| 1 | 94 (5.7%) | 247 (5.6%) | 2 (0.5%) | 671 (2.6%) | 1014 (3.1%) | |
| **conHistLdm** | | | | | | < 0.001 |
| 0 | 1632 (98.5%) | 4369 (99.6%) | 437 (100.0%) | 25712 (99.8%) | 32150 (99.7%) | |
| 1 | 25 (1.5%) | 16 (0.4%) | 0 (0.0%) | 40 (0.2%) | 81 (0.3%) | |
| **conInstit** | | | | | | < 0.001 |
| 0 | 1654 (99.8%) | 4377 (99.8%) | 437 (100.0%) | 25752 (100.0%) | 32220 (100.0%) | |
| 1 | 3 (0.2%) | 8 (0.2%) | 0 (0.0%) | 0 (0.0%) | 11 (0.0%) | |
| **conLSHA** | | | | | | < 0.001 |

|  | Mixed Use (N=1657) | Multi-family (N=4385) | Non-conforming (N=437) | Single-family (N=25752) | Total (N=32231) | p value |
|---|---|---|---|---|---|---|
| 0 | 1557 (94.0%) | 4108 (93.7%) | 399 (91.3%) | 20932 (81.3%) | 26996 (83.8%) |  |
| 1 | 100 (6.0%) | 277 (6.3%) | 38 (8.7%) | 4820 (18.7%) | 5235 (16.2%) |  |
| **conLUST** |  |  |  |  |  | < 0.001 |
| 0 | 1602 (96.7%) | 4384 (100.0%) | 407 (93.1%) | 25748 (100.0%) | 32141 (99.7%) |  |
| 1 | 55 (3.3%) | 1 (0.0%) | 30 (6.9%) | 4 (0.0%) | 90 (0.3%) |  |
| **conNoise** |  |  |  |  |  | < 0.001 |
| 0 | 1565 (94.4%) | 4291 (97.9%) | 374 (85.6%) | 25492 (99.0%) | 31722 (98.4%) |  |
| 1 | 92 (5.6%) | 94 (2.1%) | 63 (14.4%) | 260 (1.0%) | 509 (1.6%) |  |
| **conPrvCom** |  |  |  |  |  | 0.161 |
| 0 | 1656 (99.9%) | 4385 (100.0%) | 437 (100.0%) | 25750 (100.0%) | 32228 (100.0%) |  |
| 1 | 1 (0.1%) | 0 (0.0%) | 0 (0.0%) | 2 (0.0%) | 3 (0.0%) |  |
| **conSewer** |  |  |  |  |  | < 0.001 |
| 0 | 1657 (100.0%) | 4385 (100.0%) | 436 (99.8%) | 25625 (99.5%) | 32103 (99.6%) |  |
| 1 | 0 (0.0%) | 0 (0.0%) | 1 (0.2%) | 127 (0.5%) | 128 (0.4%) |  |
| **conSLIDO** |  |  |  |  |  | < 0.001 |
| 0 | 1648 (99.5%) | 4367 (99.6%) | 429 (98.2%) | 25304 (98.3%) | 31748 (98.5%) |  |
| 1 | 9 (0.5%) | 18 (0.4%) | 8 (1.8%) | 448 (1.7%) | 483 (1.5%) |  |
| **conSlp25** |  |  |  |  |  | < 0.001 |
| 0 | 1620 (97.8%) | 4269 (97.4%) | 424 (97.0%) | 23883 (92.7%) | 30196 (93.7%) |  |
| 1 | 37 (2.2%) | 116 (2.6%) | 13 (3.0%) | 1869 (7.3%) | 2035 (6.3%) |  |
| **conStorm** |  |  |  |  |  | < 0.001 |
| 0 | 1579 (95.3%) | 4140 (94.4%) | 383 (87.6%) | 23730 (92.1%) | 29832 (92.6%) |  |
| 1 | 78 (4.7%) | 245 (5.6%) | 54 (12.4%) | 2022 (7.9%) | 2399 (7.4%) |  |
| **conTranCap** |  |  |  |  |  | < 0.001 |
| 0 | 1657 (100.0%) | 4318 (98.5%) | 431 (98.6%) | 24028 (93.3%) | 30434 (94.4%) |  |
| 1 | 0 (0.0%) | 67 (1.5%) | 6 (1.4%) | 1724 (6.7%) | 1797 (5.6%) |  |
| **conTranSub** |  |  |  |  |  | < 0.001 |
| 0 | 1482 (89.4%) | 3523 (80.3%) | 318 (72.8%) | 17395 (67.5%) | 22718 (70.5%) |  |

|  | Mixed Use (N=1657) | Multi-family (N=4385) | Non-conforming (N=437) | Single-family (N=25752) | Total (N=32231) | p value |
|---|---|---|---|---|---|---|
| 1 | 175 (10.6%) | 862 (19.7%) | 119 (27.2%) | 8357 (32.5%) | 9513 (29.5%) |  |
| **conTranInt** |  |  |  |  |  | < 0.001 |
| 0 | 1655 (99.9%) | 4345 (99.1%) | 436 (99.8%) | 23677 (91.9%) | 30113 (93.4%) |  |
| 1 | 2 (0.1%) | 40 (0.9%) | 1 (0.2%) | 2075 (8.1%) | 2118 (6.6%) |  |
| **conView** |  |  |  |  |  | 0.969 |
| 0 | 1657 (100.0%) | 4385 (100.0%) | 437 (100.0%) | 25751 (100.0%) | 32230 (100.0%) |  |
| 1 | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (0.0%) | 1 (0.0%) |  |
| **conWater** |  |  |  |  |  | < 0.001 |
| 0 | 1574 (95.0%) | 4124 (94.0%) | 389 (89.0%) | 24857 (96.5%) | 30944 (96.0%) |  |
| 1 | 83 (5.0%) | 261 (6.0%) | 48 (11.0%) | 895 (3.5%) | 1287 (4.0%) |  |
| **conGW** |  |  |  |  |  | < 0.001 |
| 0 | 1653 (99.8%) | 4384 (100.0%) | 434 (99.3%) | 25730 (99.9%) | 32201 (99.9%) |  |
| 1 | 4 (0.2%) | 1 (0.0%) | 3 (0.7%) | 22 (0.1%) | 30 (0.1%) |  |
| **conPubOwn** |  |  |  |  |  | < 0.001 |
| 0 | 1651 (99.6%) | 4375 (99.8%) | 428 (97.9%) | 25734 (99.9%) | 32188 (99.9%) |  |
| 1 | 6 (0.4%) | 10 (0.2%) | 9 (2.1%) | 18 (0.1%) | 43 (0.1%) |  |
| **conFldway** |  |  |  |  |  | < 0.001 |
| 0 | 1644 (99.2%) | 4378 (99.8%) | 427 (97.7%) | 25702 (99.8%) | 32151 (99.8%) |  |
| 1 | 13 (0.8%) | 7 (0.2%) | 10 (2.3%) | 50 (0.2%) | 80 (0.2%) |  |
| **conFld100** |  |  |  |  |  | < 0.001 |
| 0 | 1632 (98.5%) | 4354 (99.3%) | 379 (86.7%) | 25524 (99.1%) | 31889 (98.9%) |  |
| 1 | 25 (1.5%) | 31 (0.7%) | 58 (13.3%) | 228 (0.9%) | 342 (1.1%) |  |
| **conECSI** |  |  |  |  |  | < 0.001 |
| 0 | 1642 (99.1%) | 4381 (99.9%) | 411 (94.1%) | 25747 (100.0%) | 32181 (99.8%) |  |
| 1 | 15 (0.9%) | 4 (0.1%) | 26 (5.9%) | 5 (0.0%) | 50 (0.2%) |  |

```
# constraints by property type table
constraints_df <- trim %>%
  select(constraints, prop_type) %>%
  mutate_all(as.factor)
const_type_tbl <- tableby(as.formula(paste("prop_type ~ ",
```

Table 2: Frequency of Sale Zones

| Var1 | Freq |
|---|---:|
| CE | 15 |
| CG | 406 |
| CM | 153 |
| CM1 | 34 |
| CM2 | 100 |
| CM3 | 20 |
| CN1 | 36 |
| CN2 | 115 |
| CO1 | 18 |
| CO2 | 24 |
| CS | 411 |
| CX | 100 |
| EG1 | 40 |
| EG2 | 77 |
| EX | 225 |
| IG1 | 122 |
| IG2 | 116 |
| IH | 80 |
| IR | 7 |
| OS | 2 |
| R1 | 1219 |
| R10 | 2176 |
| R2 | 2403 |
| R2.5 | 3128 |
| R20 | 298 |
| R3 | 412 |
| R5 | 14506 |
| R7 | 5534 |
| RF | 110 |
| RH | 311 |
| RX | 33 |

```
                                       paste(constraints, collapse = " + "))),
                                 data = constraints_df)
x <- summary(const_type_tbl, title = "Constraint frequency by property type")

# zone table
table(trim$sale_zone) %>% kable(caption = "Frequency of Sale Zones") %>%
  kable_styling("striped")
```

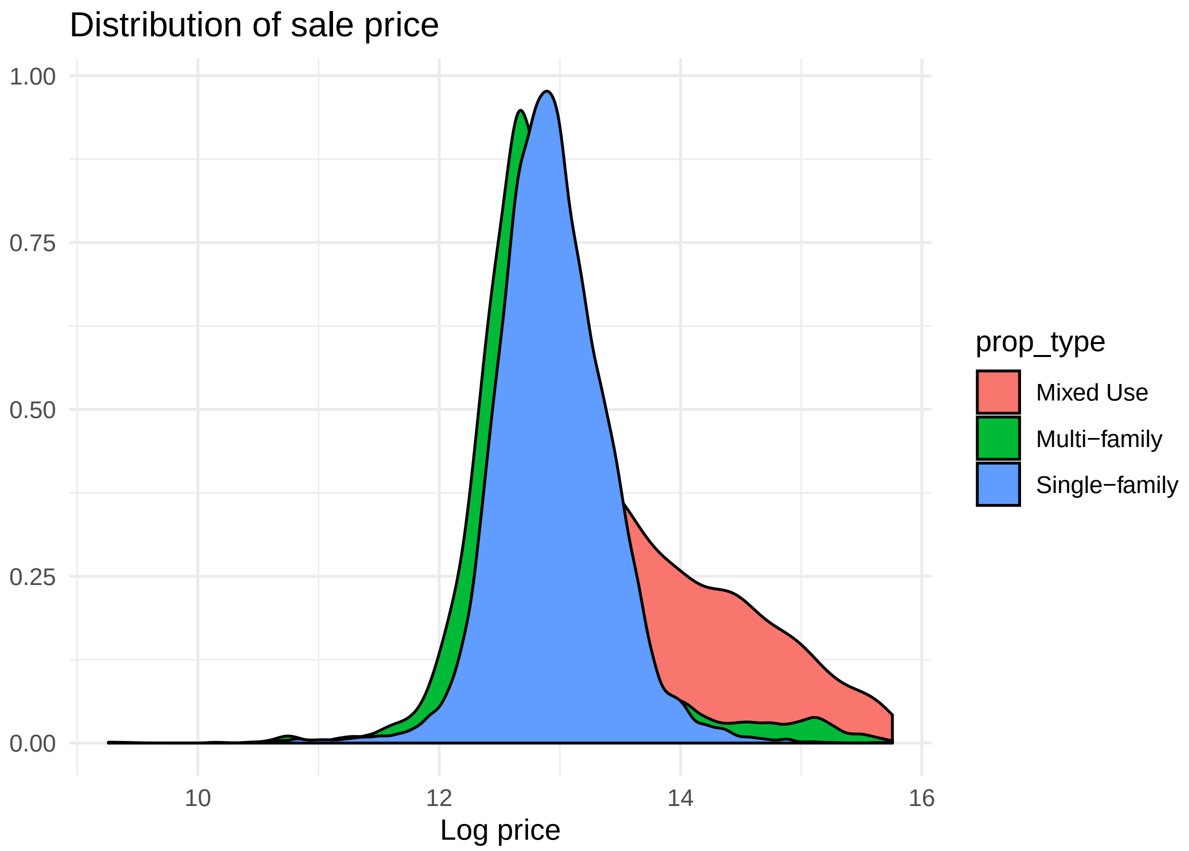
```
# property type table
table(trim$prop_type) %>%
  as.data.frame() %>%
  arrange(desc(Freq)) %>%
  kable(caption = "Frequency of Property Type") %>%
  kable_styling("striped")
```

Table 3: Frequency of Property Type

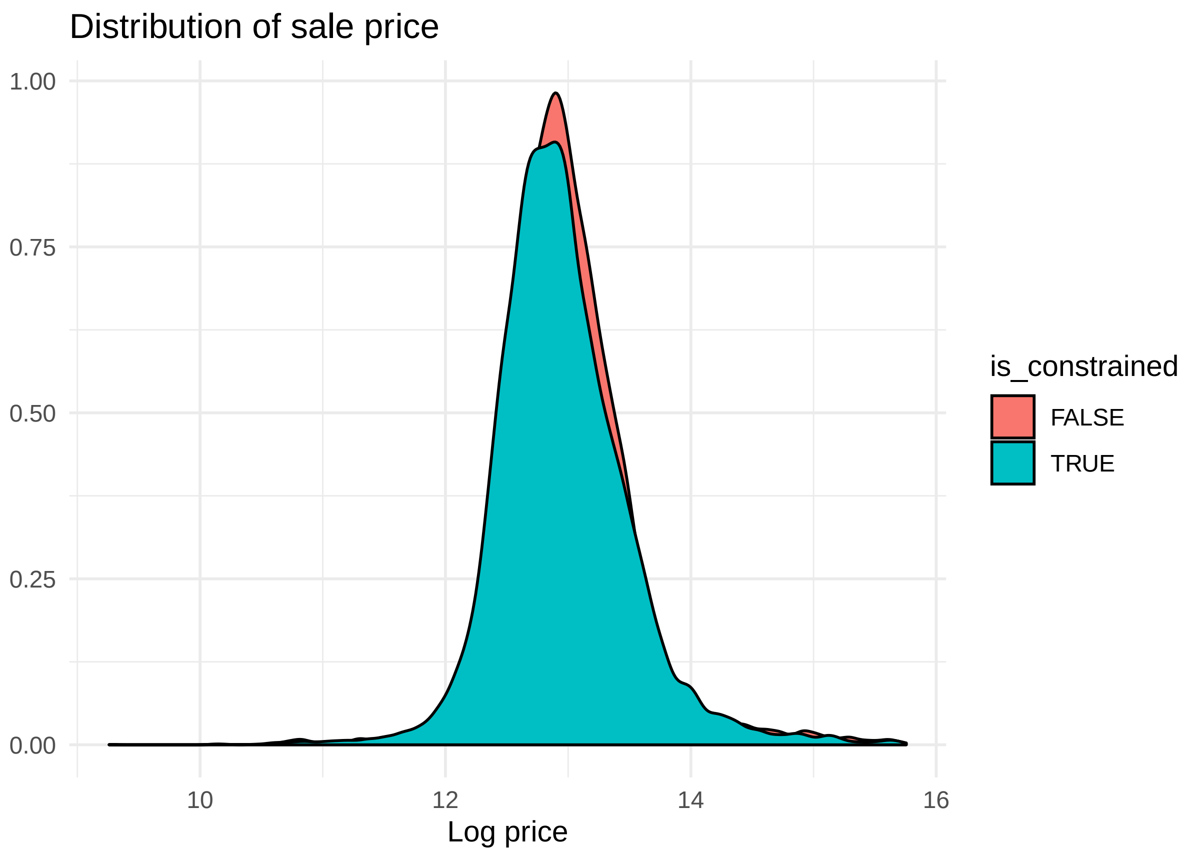
| Var1 | Freq |
|---|---|
| Single-family | 25752 |
| Multi-family | 4385 |
| Mixed Use | 1657 |
| Non-conforming | 437 |

## Plots

```r
# density plot of sale price by property type
ggplot(trim %>% filter(prop_type != "Non-conforming"),
       aes(log(SALEPRICE), fill = prop_type)) +
  geom_density() +
  labs(title = "Distribution of sale price", x = "Log price", y = "") +
  theme_minimal()
```



```r
# density plot of sale price by whether or not constrained
ggplot(trim %>% filter(prop_type != "Non-conforming"),
       aes(log(SALEPRICE), fill = is_constrained)) +
  geom_density() +
  labs(title = "Distribution of sale price", x = "Log price", y = "") +
  theme_minimal()
```



### Control variables

```
controls <- tableby(~ f_baths + h_baths + pct_canopy_cov + CN_score + dist_cityhall +
          dist_ugb + YEARBUILT, data = trim)
controls$control$numeric.stats <- c("Nmiss",
                                    "meansd",
                                    "median",
                                    "range")
#summary(controls, title = "Continuous Controls", digits = 2)
```

| | Overall (N=32231) |
|---|---|
| **f_baths** | |
| N-Miss | 1859 |
| Mean (SD) | 1.864 (1.180) |
| Range | 0.000 - 40.000 |
| **h_baths** | |
| N-Miss | 1866 |
| Mean (SD) | 0.351 (0.590) |
| Range | 0.000 - 20.000 |
| **pct_canopy_cov** | |
| N-Miss | 1101 |
| Mean (SD) | 0.282 (0.198) |
| Range | 0.000 - 1.000 |

| | Overall (N=32231) |
|---|---|
| **CN__score** | |
| Mean (SD) | 58.094 (17.443) |
| Range | 1.000 - 100.000 |
| **dist__cityhall** | |
| Mean (SD) | 25274.303 (10161.726) |
| Range | 1122.948 - 53006.519 |
| **dist__ugb** | |
| Mean (SD) | 18713.615 (7543.791) |
| Range | 17.204 - 34315.313 |
| **YEARBUILT** | |
| N-Miss | 114 |
| Mean (SD) | 1948.780 (335.074) |
| Range | 0.000 - 9999.000 |