

Where should you live in San Francisco?

Maximizing housing value in the world's most expensive city with Zillow data

By Collin Ching

```
library(knitr)
```

```
opts_chunk$set(warning=FALSE,message=FALSE)
```

There's plenty to love about San Francisco... except for the cost of living. Housing is a big chunk of that, so it's important to get the most value out of your housing situation whether you're a renter or homeowner.

In this project, I explore Zillow data in R to identify high-value neighborhoods for 1) prospective renters and 2) investors looking to buy rental property. We'll explore home prices, monthly rent, and a ratio of the two-price-to-rent-to find candidate neighborhoods.

Thanks to Ken Steif and Keith Hassel for their great tutorial on exploring home prices in San Francisco. They inspired the following plots and graphics.

```
## load libraries and local files
options(scipen="99")
```

```
library(cowplot)
library(data.table)
library(broom)
library(tidyverse)
library(ggmap)
library(ggrepel)
library(grid)
library(gridExtra)
library(mapttools)
library(scales)
library(sf)
source("../workflow/functions.R")
```

```
## smoothed median home price dataset
zhvi <- read.csv("../data/zillow/Neighborhood_Zhvi_AllHomes.csv",stringsAsFactors=FALSE)
## smoothed median neighborhood rent dataset
zri <- read.csv("../data/zillow/Neighborhood_Zri_AllHomesPlusMultifamily.csv",stringsAsFactors=FALSE)
## smoothed median rent per square foot dataset
zri_sqft <- read.csv("../data/zillow/Neighborhood_ZriPerSqft_AllHomes.csv",stringsAsFactors=FALSE)
## ratio of estimated home value to 12*monthly rent, renters would want this to be smaller
price_to_rent <- read.csv("../data/zillow/Neighborhood_PriceToRentRatio_AllHomes.csv",stringsAsFactors=FALSE)
## sales-to-list ratio: tells you if market is favoring buyers
## time-on-market: the amount of time that homes are selling on the market

## shapefile of SF neighborhoods
sf_neighborhoods <- readShapePoly("../sf_neighborhoods/sf_neighborhoods")
```

Data processing

```
## tidy data
zhvi.sf <- zhvi[zhvi$City=="San Francisco",]
zhvi.sf <- gather(zhvi.sf,date,zhvi,8:ncol(zhvi.sf)) %>%
  mutate(date=str_sub(date,start=2),
         year=str_sub(date,1,4),
         month=str_sub(date,-2,-1)) %>%
  rename(neighborhood=RegionName) %>%
  select(neighborhood,year,month,zhvi)

zri.sf <- zri[zri$City=="San Francisco",]
zri.sf <- zri.sf %>%
  gather(date,zri,8:ncol(zri.sf)) %>%
  mutate(date=str_sub(date,start=2),
         year=str_sub(date,1,4),
         month=str_sub(date,-2,-1)) %>%
  rename(neighborhood=RegionName,city=City) %>%
  select(neighborhood,year,month,zri)

zri_sqft.sf <- zri_sqft[zri_sqft$City=="San Francisco",]
zri_sqft.sf <- zri_sqft.sf %>%
  gather(date,zri_sqft,8:ncol(zri_sqft)) %>%
  mutate(date=str_sub(date,start=2),
         year=str_sub(date,1,4),
         month=str_sub(date,-2,-1)) %>%
  rename(neighborhood=RegionName) %>%
  select(neighborhood,year,month,zri_sqft)

price_to_rent.sf <- price_to_rent[price_to_rent$City=="San Francisco",]
price_to_rent.sf <- price_to_rent.sf %>%
  gather(date,price_to_rent,8:ncol(price_to_rent.sf)) %>%
  mutate(date=str_sub(date,start=2),
         year=str_sub(date,1,4),
         month=str_sub(date,-2,-1)) %>%
  rename(neighborhood=RegionName) %>%
  select(neighborhood,year,month,price_to_rent)

#names(zhvi.sf)
#names(zri.sf)
#names(zri_sqft.sf)
#names(price_to_rent.sf)

sf_housing <- left_join(zhvi.sf,zri.sf,by=c("neighborhood","year","month")) %>%
  left_join(.,zri_sqft.sf,by=c("neighborhood","year","month")) %>%
  left_join(.,price_to_rent.sf,by=c("neighborhood","year","month"))

sf_housing$year <- as.integer(sf_housing$year)
sf_housing$month <- as.integer(sf_housing$month)

#####
# Make points dataset for geom_polygon()
# Rename neighborhoods in Zillow data to match shapefiles
```

```

## extract neighborhood names from spatial object
neighb_pts <- tidy(sf_neighborhoods,region="nbrhood") ## converts to a data frame
neighb_pts <- neighb_pts %>%
  rename(neighborhood=id) %>%
  select(long,lat,neighborhood)
names(neighb_pts)[1:2] <- c("lat","long") ## lat and long were switched in original data

## check unique neighborhoods in sf_housing
#length(unique(sf_housing$neighborhood))
## these are the neighborhoods in sf_housing that don't show up in neighborhood shapefile
#setdiff(unique(sf_housing$neighborhood),unique(neighb_pts$neighborhood))

## match neighborhood names between datasets
#length(setdiff(unique(neighb_pts$neighborhood),unique(sf_housing$neighborhood)))
## 61 neighborhoods in ZHVI dataset, with 11 neighborhoods mismatched
sf_housing$neighborhood[sf_housing$neighborhood=="Buena Vista"] <- "Buena Vista Park/Ashbury Heights"
sf_housing$neighborhood[sf_housing$neighborhood=="Financial District"] <- "Financial District/Barbary Coast"
sf_housing$neighborhood[sf_housing$neighborhood=="Haight"] <- "Haight Ashbury"
sf_housing$neighborhood[sf_housing$neighborhood=="Lake"] <- "Lake Street"
sf_housing$neighborhood[sf_housing$neighborhood=="Lakeshore"] <- "Lake Shore"
sf_housing$neighborhood[sf_housing$neighborhood=="Laurel Heights"] <- "Jordan Park / Laurel Heights"
sf_housing$neighborhood[sf_housing$neighborhood=="Mission"] <- "Mission Dolores"
sf_housing$neighborhood[sf_housing$neighborhood=="Panhandle"] <- "North Panhandle"
sf_housing$neighborhood[sf_housing$neighborhood=="Seacliff"] <- "Sea Cliff"
sf_housing$neighborhood[sf_housing$neighborhood=="St. Francis Wood"] <- "Saint Francis Wood"
sf_housing$neighborhood[sf_housing$neighborhood=="Upper Market"] <- "Twin Peaks"
## still some missing, but we're limited to our data

#####
## Add districts to dataset
presidio <- c("Presidio",
             "Sea Cliff",
             "Presidio Heights",
             "Pacific Heights",
             "Cow Hollow",
             "Marina",
             "Lake Street")

nob_hill <- c("Russian Hill",
             "North Beach",
             "Nob Hill",
             "Telegraph Hill",
             "North Waterfront",
             "Financial District/Barbary Coast",
             "Downtown")

avenues <- c("Lincoln Park",
            "Outer Richmond",
            "Central Richmond",
            "Inner Richmond",
            "Golden Gate Park",
            "Outer Sunset",
            "Central Sunset",

```

```

    "Outer Parkside",
    "Parkside",
    "Pine Lake Park",
    "Inner Sunset")

twinpeaks <- c("Lake Shore",
    "Stonestown",
    "Merced Manor",
    "Lakeside",
    "Ingleside Terrace",
    "Merced Heights",
    "Balboa Terrace",
    "Mount Davidson Manor",
    "Westwood Park",
    "Westwood Highlands",
    "Sunnyside",
    "Miraloma Park",
    "Diamond Heights",
    "Twin Peaks",
    "Midtown Terrace",
    "Golden Gate Heights",
    "Forest Knolls",
    "Forest Hill",
    "Forest Hills Extension",
    "West Portal",
    "Saint Francis Wood",
    "Sherwood Forest",
    "Monterey Heights",
    "Mount Davidson Manor",
    "Inner Parkside",
    "Clarendon Heights")

castro <- c("Noe Valley",
    "Eureka Valley / Dolores Heights",
    "Corona Heights",
    "Diamond Heights",
    "Glen Park",
    "Duboce Triangle")

mission <- c("Bernal Heights", "Inner Mission", "Mission Dolores")

southeast <- c("Mission Terrace",
    "Ingleside",
    "Ingleside Heights",
    "Outer Mission",
    "Oceanview",
    "Crocker Amazon",
    "Excelsior",
    "Portola",
    "Silver Terrace",
    "Bayview",
    "Bayview Heights",
    "Visitation Valley",

```

```

    "Little Hollywood",
    "Candlestick Point",
    "Hunters Point",
    "Central Waterfront/Dogpatch",
    "Potrero Hill",
    "Excelsior")

soma <- c("Mission Bay","South of Market","South Beach","Yerba Buena")

haight <- c("Haight Ashbury",
            "Buena Vista Park/Ashbury Heights",
            "Cole Valley/Parnassus Heights")

tenderloin <- c("Tenderloin","Van Ness/Civic Center")

western_addition <- c("Anza Vista",
                     "Lower Pacific Heights",
                     "Hayes Valley",
                     "Alamo Square",
                     "Western Addition",
                     "North Panhandle",
                     "Jordan Park / Laurel Heights",
                     "Lone Mountain")

neighb_pts$district[neighb_pts$neighborhood %in% western_addition] <- "Western Addition"
neighb_pts$district[neighb_pts$neighborhood %in% avenues] <- "The Avenues"
neighb_pts$district[neighb_pts$neighborhood %in% twinpeaks] <- "Twin Peaks"
neighb_pts$district[neighb_pts$neighborhood %in% presidio] <- "The Presidio"
neighb_pts$district[neighb_pts$neighborhood %in% nob_hill] <- "Nob Hill/FiDi/Downtown"
neighb_pts$district[neighb_pts$neighborhood %in% castro] <- "Castro"
neighb_pts$district[neighb_pts$neighborhood %in% mission] <- "Mission"
neighb_pts$district[neighb_pts$neighborhood %in% southeast] <- "Southeast San Francisco"
neighb_pts$district[neighb_pts$neighborhood %in% soma] <- "SoMa"
neighb_pts$district[neighb_pts$neighborhood %in% haight] <- "Haight-Ashbury"
neighb_pts$district[neighb_pts$neighborhood %in% tenderloin] <- "Civic Center/Tenderloin"

#####
## Aggregate sf_housing by year
sf_housing.yearly <- sf_housing %>%
  group_by(neighborhood,year) %>%
  summarize(zhvi=mean(zhvi,na.rm=TRUE),
            zri=mean(zri,na.rm=TRUE),
            zri_sqft=mean(zri_sqft,na.rm=TRUE),
            price_to_rent=mean(price_to_rent,na.rm=TRUE))

#names(sf_housing)
#names(sf_housing.yearly)

#####
## Join neighborhood points with descriptive statistics
#names(sf_housing)
#names(neighb_pts)
sf_housing_pts <- right_join(neighb_pts,sf_housing,by="neighborhood")

```

```

sf_housing_pts.yearly <- right_join(neighb_pts,sf_housing.yearly,by="neighborhood")
#str(sf_housing_pts)
#str(sf_housing_pts.yearly)

#####
## Load map objects
## get coordinates for map boundary
bbox <- sf_neighborhoods@bbox
sf_bbox <- c(left=bbox[1,1]-.01,
             bottom=bbox[2,1]-.005,
             right=bbox[1,2]+.01,
             top=bbox[2,2]+.005)

## load basemap from Stamen maps
basemap <- get_stamenmap(
  bbox = sf_bbox,
  zoom=12,
  maptype="toner-lite")

```

Neighborhood reference

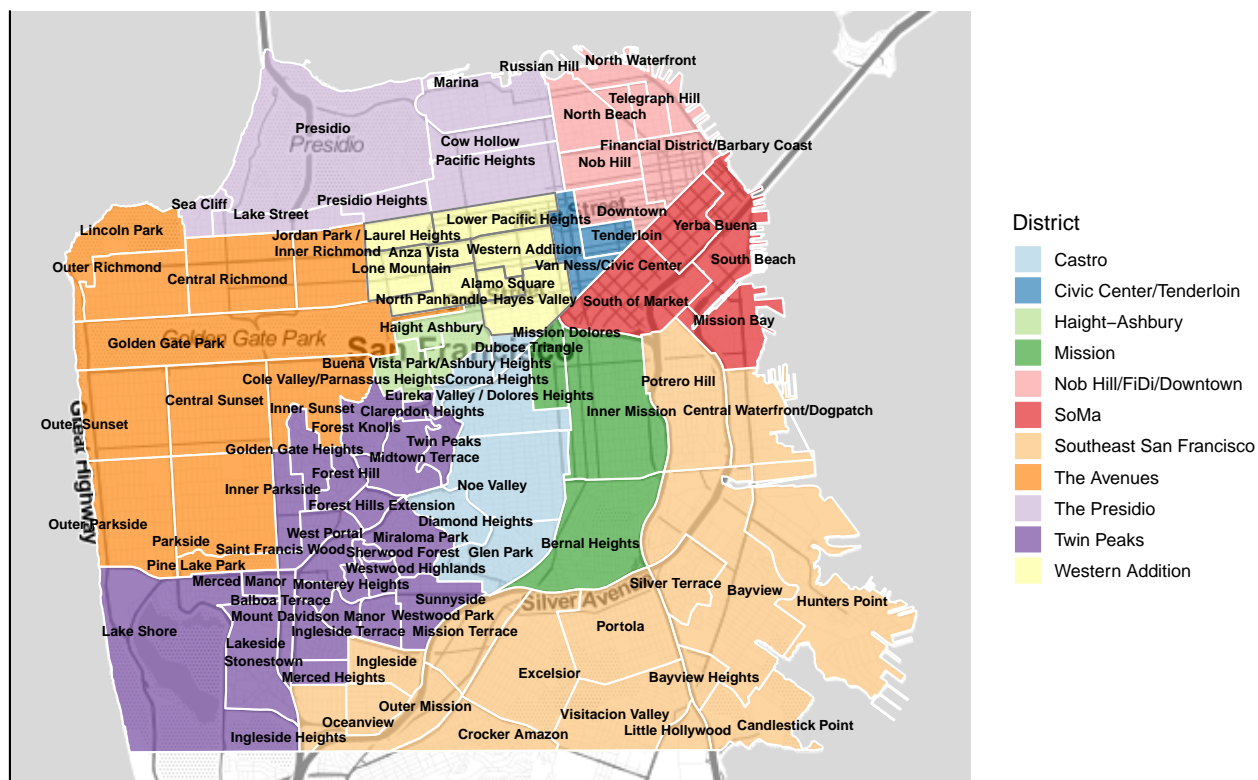
Since my analysis is done with respect to neighborhoods, here's a visual reference.

```

neighb_centroids <- neighb_pts %>%
  group_by(neighborhood) %>%
  summarize(mdn_lat=median(lat,na.rm=TRUE),
            mdn_long=median(long,na.rm=TRUE),
            mean_lat=mean(lat,na.rm=TRUE),
            mean_long=mean(long,na.rm=TRUE)) %>%
  as.data.frame()

ggmap(basemap) + coord_map() +
  mapTheme() +
  geom_polygon(data=neighb_pts,
              aes(x=lat,y=long,group=neighborhood,fill=neighborhood),
              col="white",alpha=.65) +
  geom_polygon(data=filter(neighb_pts,district=="Western Addition"),
              aes(x=lat,y=long,group=neighborhood),
              col="gray50",alpha=0,show.legend=FALSE) +
  ## ggrepel library prevents text overlapping
  geom_text_repel(data=neighb_centroids,
                  aes(x=mean_lat,y=mean_long,label=neighborhood),
                  size=2.34,color="black",fontface="bold",
                  segment.size=.25,point.padding=NA,box.padding=.1,force=.1) +
  scale_fill_brewer("District",palette="Paired",type="qual")

```



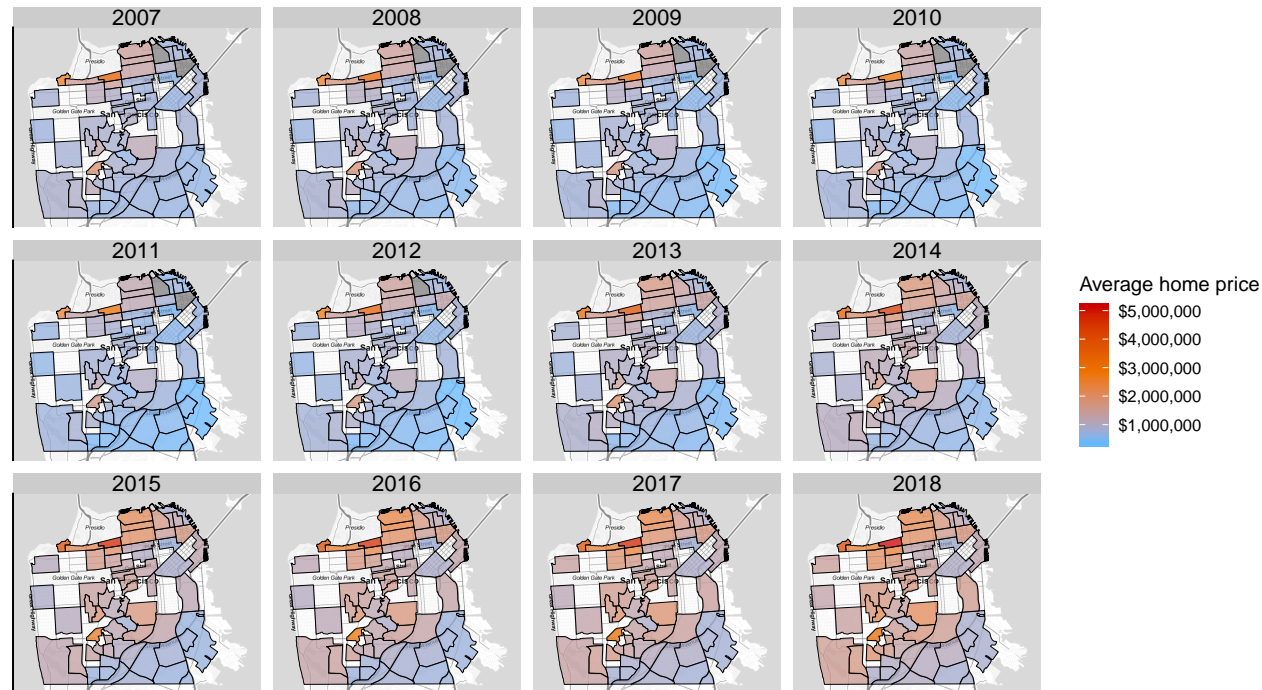
```
#ggsave("../graphics/fig1.png",width=9,height=9)
```

San Francisco home prices by year

```
ggmap(basemap) + coord_map() +
  geom_polygon(data=filter(sf_housing_pts.yearly,year>=2007),
    aes(x=lat,y=long,group=neighborhood,fill=zhvi),
    col="black",alpha=.75,size=.25) +
  facet_wrap(~year,nrow=3) +
  labs(title="San Francisco home prices",
    subtitle="Mean Zillow Home Value Indices") +
  scale_fill_gradientn("Average home price",
    colors=c("steelblue1","darkorange2","red3"),
    labels = scales::dollar_format(prefix = "$")) +
  mapTheme()
```


San Francisco home prices

Mean Zillow Home Value Indices



```
#ggsave("../graphics/fig2.png",width=10,height=6)
```

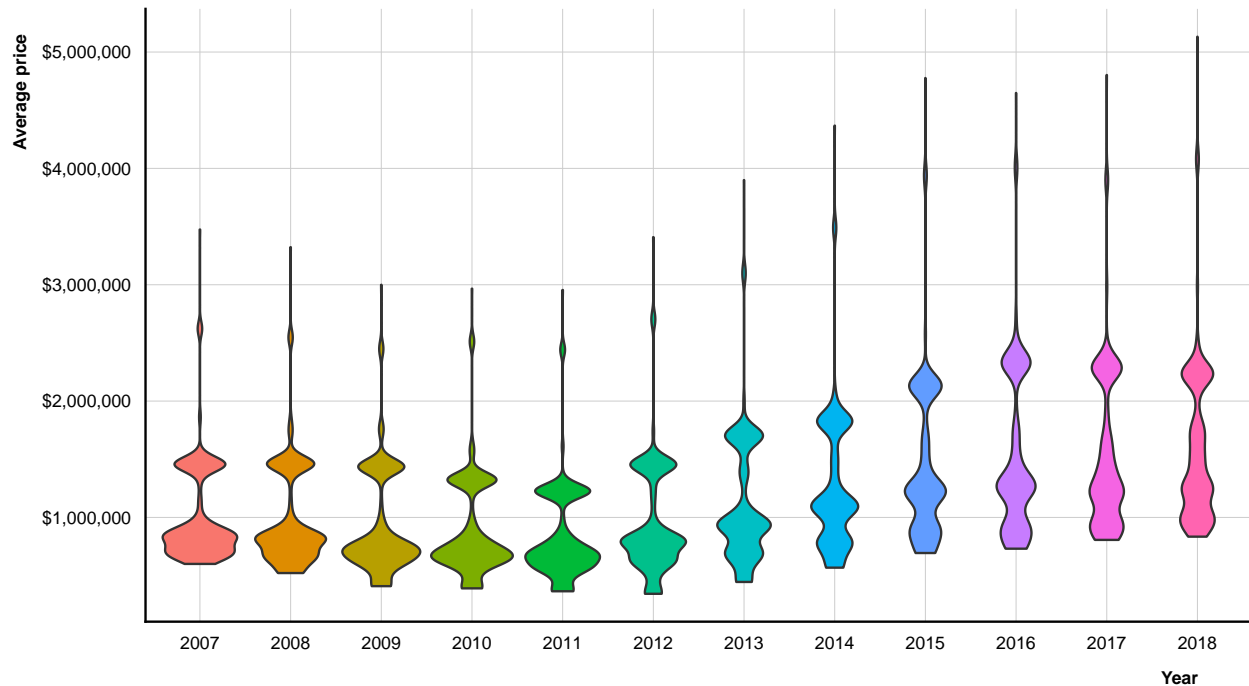
Neighborhood home prices are stagnant between 2007 and 2011 due to the subprime mortgage crisis. In the Shipyard, housing prices even decrease during this time period.

From 2012 onwards, housing prices take off. This rise in housing grows in the center of San Francisco, then propagates outwards. Housing is consistently most expensive near the Presidio and Twin Peaks areas.

```
ggplot(filter(sf_housing_pts.yearly,year>=2007)) +
  plotTheme() + theme(legend.position="none") +
  geom_violin(aes(x=as.factor(year),y=zghi,fill=as.factor(year)),size=.5) +
  labs(title="Distribution of average neighborhood home prices",
        subtitle="Mean Zillow Home Value Indices",
        x="Year",
        y="Average price") +
  scale_y_continuous(labels = dollar_format(prefix = "$"))
```


Distribution of average neighborhood home prices

Mean Zillow Home Value Indices



```
#ggsave("graphics/fig3.png",width=8,height=5,device="png")
```

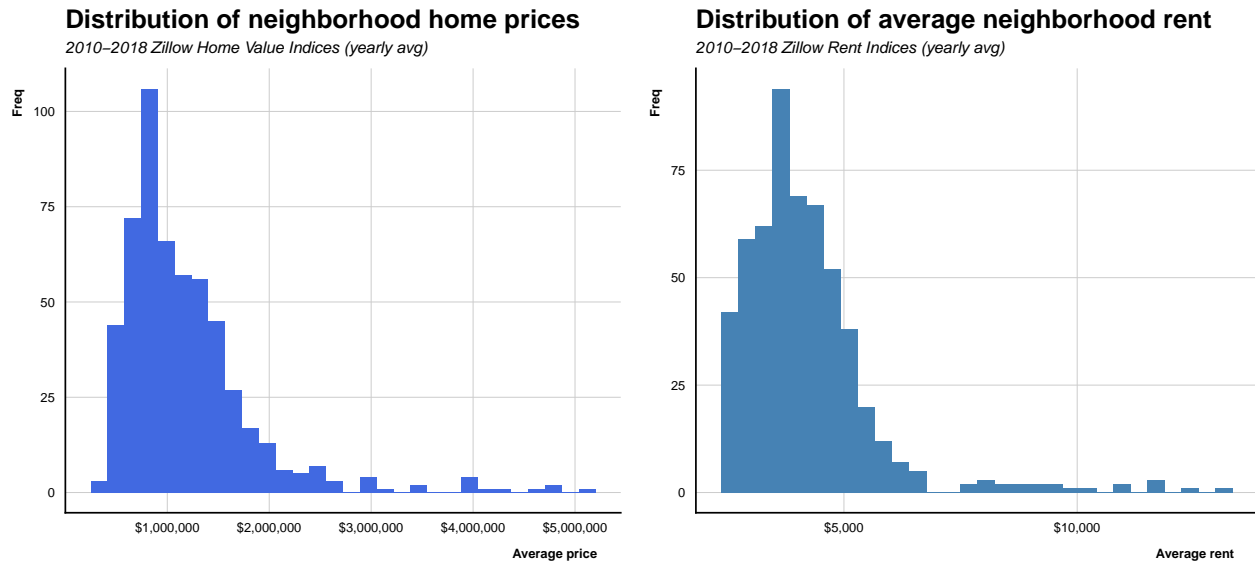
This is a better view of change over time. Housing prices decrease during the subprime mortgage crisis, then shoot up after 2012 and split into a trimodal distribution.

Comparing home prices and rent

```
prices <- ggplot(filter(sf_housing.yearly,year>=2010),aes(x=zhvi)) +
  geom_histogram(fill="royalblue") +
  plotTheme() +
  scale_x_continuous(labels=dollar_format(prefix="$")) +
  labs(title="Distribution of neighborhood home prices",
       subtitle="2010-2018 Zillow Home Value Indices (yearly avg)",
       x="Average price",
       y="Freq")

rent <- ggplot(filter(sf_housing.yearly,year>=2010),aes(x=zri)) +
  geom_histogram(fill="steelblue") +
  plotTheme() +
  scale_x_continuous(labels=dollar_format(prefix="$")) +
  labs(title="Distribution of average neighborhood rent",
       subtitle="2010-2018 Zillow Rent Indices (yearly avg)",
       x="Average rent",
       y="Freq")

grid.arrange(prices,rent,ncol=2)
```



```
#ggsave("../graphics/4.png",width=11,height=5)
```

The plots above includes all home types—condos, single-family, multi-family—and home sizes—bedroom, 2-bedroom, 3-bedroom, etc.

The typical home price was a little under \$1 million and typical rent was roughly \$4,000 per month from 2010 to 2018. There are also rare neighborhoods where the median home price exceeds \$3,000,000 and median rent exceeds \$6,000.

Home price and rent follow a similar distribution.

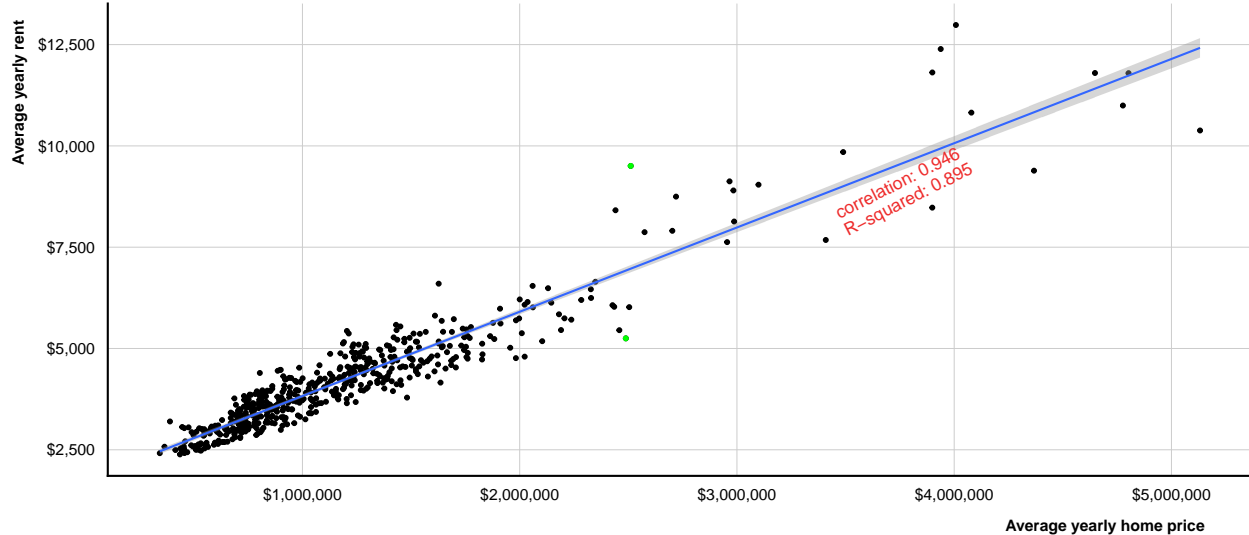
```
cor.zhvi_zri <- cor(sf_housing.yearly$zhvi[sf_housing.yearly$year>=2007],
  sf_housing.yearly$zri[sf_housing.yearly$year>=2007],
  use = "complete.obs") %>%
  round(3)

fit.zhvi_zri <- lm(zri~zhvi,filter(sf_housing.yearly,year>=2007))

ggplot(filter(sf_housing.yearly,year>=2007),aes(x=zhvi,y=zri)) +
  geom_point(size=.75) +
  geom_smooth(method="lm",size=.5) +
  geom_point(data=sf_housing.yearly[c(506,1142),],aes(x=zhvi,y=zri),col="green",size=.75) +
  labs(title="San Francisco rent vs home prices",
    subtitle="2010–2018 neighborhood Zillow Home Value Indices and Zillow Rent Indices",
    x="Average yearly home price",
    y="Average yearly rent") +
  plotTheme() +
  scale_x_continuous(labels=dollar_format(prefix="$")) +
  scale_y_continuous(labels=dollar_format(prefix="$")) +
  annotate("text",x=3400000,y=8940,
    label=paste(paste("correlation:",cor.zhvi_zri)),
    angle=27,size=3,col="firebrick2",vjust=2.7,hjust=0) +
  annotate("text",x=3400000,y=8940,
    label=paste(paste("R-squared:",round(summary(fit.zhvi_zri)$r.squared,3))),
    angle=27,size=3,col="firebrick2",vjust=4.2,hjust=0)
```

San Francisco rent vs home prices

2010–2018 neighborhood Zillow Home Value Indices and Zillow Rent Indices



```
#ggsave("../graphics/fig5.png",width=9,height=6)
```

Average yearly home price explains 89.5% of the variation in average yearly rent. Still, rent can vary at a fixed home price: at \$2.5m, the minimum rent is \$5250/mo and maximum is \$9506/mo.

If someone offered me two \$2.5m homes to rent—one at \$5250/mo and one at \$9506/mo—I'd take the \$5250/mo most cases.

In the scenario above, I fixed home price. Say you fix your rent budget instead. **What is the highest home value you could get for that budget?** If you're purchasing rental properties, **what's the highest monthly rent you could get for a fixed purchasing budget?** These questions launch us into the next analysis.

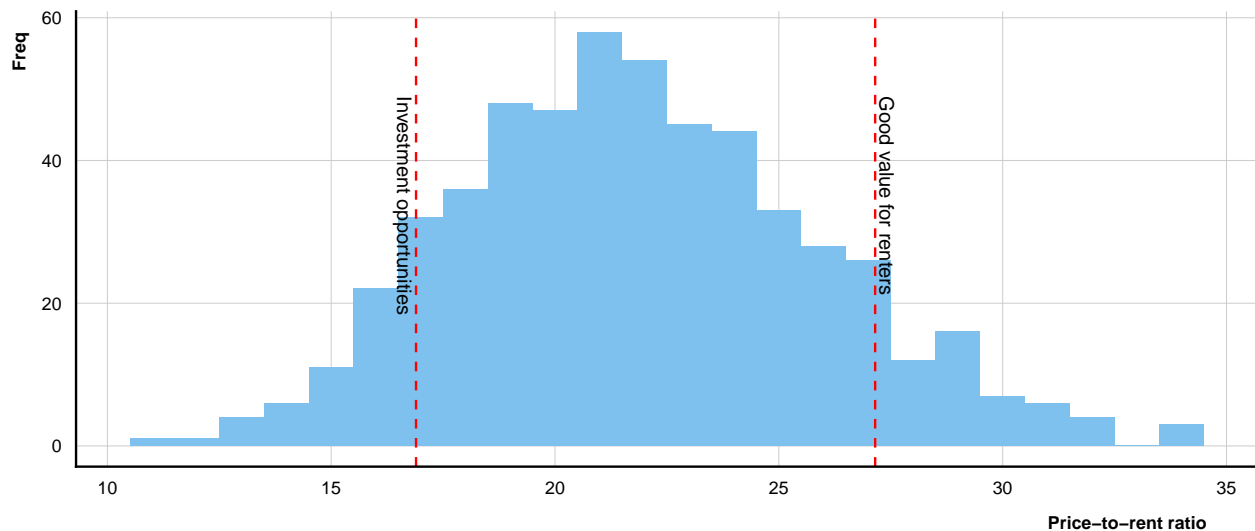
Price-to-rent

```
ptr_deciles <- quantile(sf_housing.yearly$price_to_rent[sf_housing.yearly$year>=2010],
                        seq(0,1,.1),
                        na.rm=TRUE)
```

```
ggplot(filter(sf_housing.yearly,year>=2010),aes(x=round(price_to_rent,1))) +
  geom_histogram(fill="skyblue2",binwidth=1) +
  geom_vline(aes(xintercept=ptr_deciles[2]),linetype="dashed",col="red") +
  annotate("text",x=ptr_deciles[2],y=49,angle=270,label="Investment opportunities",vjust=1.3,hjust=0,s) +
  annotate("text",x=ptr_deciles[10],y=49,angle=270,label="Good value for renters",vjust=-.3,hjust=0,s) +
  geom_vline(aes(xintercept=ptr_deciles[10]),linetype="dashed",col="red") +
  plotTheme() +
  labs(title="Distribution of neighborhood home price-to-rent ratio",
       subtitle="2010-2018 Zillow home price and rent estimates",
       x="Price-to-rent ratio",
       y="Freq") +
  theme(legend.position="none")
```

Distribution of neighborhood home price-to-rent ratio

2010–2018 Zillow home price and rent estimates



```
#ggsave("../graphics/fig6.png",width=8,height=4)
```

These are average neighborhood price-to-rent ratios from 2010 to 2018, where price-to-rent ratio is

$$\frac{\text{home price}}{12 \times \text{monthly rent}}$$

Price-to-rent is home price divided by a year's estimated rent. You can view it as the estimated years to payoff rental property. Maximizing this ratio would give renters good value; minimizing this ratio (fewer years to payoff) would give landlords good returns.

If you're renting an apartment, you want to maximize price-to-rent. These neighborhoods have higher bang for rental buck. Highest 10% neighborhoods are marked in the right panel.

If you're investing in rental property, you want to minimize price-to-rent so you get quick returns on investment. Lowest 10% price-to-rent neighborhoods are marked in the left.

```
ptr_quartiles <- quantile(sf_housing.yearly$price_to_rent[sf_housing.yearly$year>=2010],
  seq(0,1,.25),
  na.rm=TRUE)

high_ptr.10 <- filter(sf_housing.yearly,year>=2010,median(price_to_rent,na.rm=TRUE)>=ptr_deciles[10])
high_ptr_neighbs.10 <- unique(high_ptr.10$neighborhood)

high_ptr.25 <- filter(sf_housing.yearly,year>=2010,median(price_to_rent,na.rm=TRUE)>=ptr_quartiles[4])
high_ptr.25$pctile <- "75th percentile"
high_ptr.25$pctile[high_ptr.25$neighborhood %in% high_ptr_neighbs.10] <- "90th percentile"

high_ptr.25_agg <- high_ptr.25 %>%
  group_by(neighborhood) %>%
  summarize(mdn.zhvi=median(zhvi,na.rm=TRUE),
    mdn.zri=median(zri,na.rm=TRUE),
    mdn.price_to_rent=median(price_to_rent,na.rm=TRUE))

high_ptr.25 <- left_join(high_ptr.25,high_ptr.25_agg,by="neighborhood")
high_ptr_pts.25 <- right_join(neighb_pts,high_ptr.25,"neighborhood")
```

```

high_ptr.boxplot <- ggplot(high_ptr.25) +
  geom_boxplot(aes(x=reorder(factor(neighborhood),
                                price_to_rent,
                                FUN=function(x) median(x,na.rm=TRUE)),
                                y=price_to_rent,
                                fill=mdn.price_to_rent,
                                col=pctile)) +
  plotTheme() + coord_flip() +
  labs(title="Neighborhoods with highest median price-to-rent ratio",
        subtitle="2010-2018 yearly median price-to-rent (75th percentile)",
        x="Neighborhood",
        y="Price-to-rent") +
  theme(legend.position="none") +
  scale_y_continuous(breaks=c(20,22:30)) +
  scale_fill_gradientn(colors=c("lightgoldenrodyellow","forestgreen")) +
  scale_color_manual(values=c("black","deeppink")) +
  geom_vline(xintercept=9.5,col="deeppink",linetype="dotted") +
  annotate("text",x=9.5,y=20,label="Top 10%",vjust=-.5,hjust=.2,size=2.4)

## for plotting statistics on map
high_ptr_centroids <- right_join(neighb_centroids,high_ptr.25,by="neighborhood")
## adjust some points
high_ptr_centroids$mean_lat[high_ptr_centroids$neighborhood=="Saint Francis Wood"] <-
  high_ptr_centroids$mean_lat[high_ptr_centroids$neighborhood=="Saint Francis Wood"] -.003
high_ptr_centroids$mean_long[high_ptr_centroids$neighborhood=="Presidio Heights"] <-
  high_ptr_centroids$mean_long[high_ptr_centroids$neighborhood=="Presidio Heights"] + .0001
high_ptr_centroids$mean_lat[high_ptr_centroids$neighborhood=="Presidio Heights"] <-
  high_ptr_centroids$mean_lat[high_ptr_centroids$neighborhood=="Presidio Heights"] + .001
high_ptr_centroids$mean_long[high_ptr_centroids$neighborhood=="Jordan Park / Laurel Heights"] <-
  high_ptr_centroids$mean_long[high_ptr_centroids$neighborhood=="Jordan Park / Laurel Heights"] + .00
high_ptr_centroids$mean_long[high_ptr_centroids$neighborhood=="Lake Street"] <-
  high_ptr_centroids$mean_long[high_ptr_centroids$neighborhood=="Lake Street"] -.0006
high_ptr_centroids$mean_lat[high_ptr_centroids$neighborhood=="Forest Hill"] <-
  high_ptr_centroids$mean_lat[high_ptr_centroids$neighborhood=="Forest Hill"] +.0007

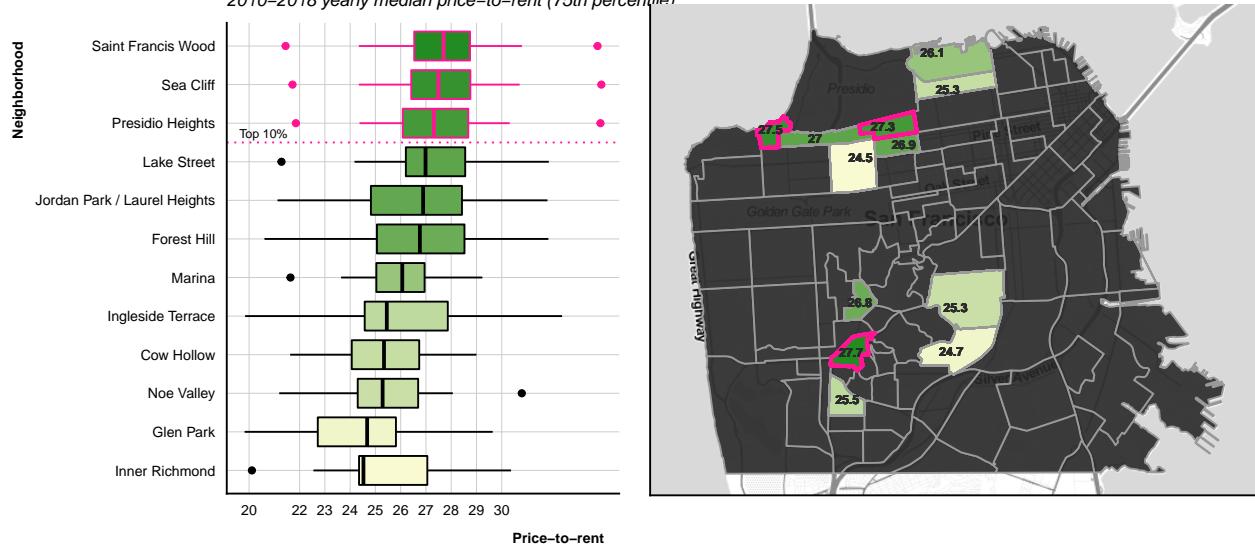
high_ptr.map <- ggmap(basemap) + coord_map() +
  geom_polygon(data=neighb_pts,
               aes(x=lat,y=long,group=neighborhood),
               col="gray60",fill="gray15",alpha=.9) +
  geom_polygon(data=filter(high_ptr_pts.25,pctile=="75th percentile"),
               aes(x=lat,y=long,group=neighborhood,fill=mdn.price_to_rent),
               col="gray60") +
  geom_polygon(data=filter(high_ptr_pts.25,pctile=="90th percentile"),
               aes(x=lat,y=long,group=neighborhood,fill=mdn.price_to_rent),
               col="deeppink",size=1) +
  geom_text(data=high_ptr_centroids,
            aes(x=mean_lat,y=mean_long,label=round(mdn.price_to_rent,1)),
            size=2.5,color="gray17",fontface="bold",nudge_y=-.00065) +
  mapTheme() + theme(legend.position="none") +
  scale_fill_gradientn(colors=c("lightgoldenrodyellow","forestgreen"))

grid.arrange(high_ptr.boxplot,high_ptr.map,nrow=1)

```

Neighborhoods with highest median price-to-rent ratio

2010–2018 yearly median price-to-rent (75th percentile)



```
#ggsave("../graphics/fig6.png",fig6,width=10,height=4.85)
```

Saint Francis Wood, Sea Cliff, and Presidio Heights are high value neighborhoods to rent in. Because they're affluent neighborhoods, home values are likely to be higher in those neighborhoods. If you can find reasonable rent in these areas, you're probably getting high value from this renting situation.

As a sanity check, I found nice Craigslist listings in Saint Francis Wood and Presidio for about \$1500/br, and the apartments looked nice. That's super reasonable for the area.

```
low_ptr.10 <- filter(sf_housing.yearly,year>=2010,median(price_to_rent,na.rm=TRUE)<=ptr_deciles[2])
low_ptr_neighbs.10 <- unique(low_ptr.10$neighborhood)

low_ptr.25 <- filter(sf_housing.yearly,year>=2010,median(price_to_rent,na.rm=TRUE)<=ptr_quartiles[2])
low_ptr.25$pctile <- "Max 25th percentile"
low_ptr.25$pctile[low_ptr.25$neighborhood %in% low_ptr_neighbs.10] <- "Max 10th percentile"

## create aggregates
low_ptr.25_agg <- low_ptr.25 %>%
  group_by(neighborhood) %>%
  summarize(mdn.zhvi=median(zhvi,na.rm=TRUE),
            mdn.zri=median(zri,na.rm=TRUE),
            mdn.price_to_rent=median(price_to_rent,na.rm=TRUE))

low_ptr.25 <- left_join(low_ptr.25,low_ptr.25_agg,by="neighborhood")
low_ptr_pts.25 <- right_join(neighb_pts,low_ptr.25,"neighborhood")

low_ptr.boxplot <- ggplot(low_ptr.25) +
  geom_boxplot(aes(x=reorder(factor(neighborhood),
                             price_to_rent,
                             FUN=function(x) median(x,na.rm=TRUE)),
                  y=price_to_rent,
                  fill=mdn.price_to_rent,,
                  color=pctile)) +
  plotTheme() + coord_flip() +
  labs(title="Neighborhoods with lowest median price-to-rent",
       subtitle="2010-2018 median yearly price-to-rent ratio (Bottom 25%)",
```

```

    x="Neighborhood",
    y="Price-to-rent") +
  scale_fill_gradientn(colors=c("forestgreen","lightgoldenrodyellow")) +
  scale_y_continuous(breaks=c(10,15:20)) +
  theme(legend.position="none") +
  geom_vline(xintercept=5.5,col="deeppink",linetype="dotted") +
  annotate("text",x=5.5,y=11,vjust=1.5,hjust=.3,label="Bottom 10%",size=2.3) +
  scale_color_manual(values=c("deeppink","black"))

## names(sf_housing_pts.yearly)
low_ptr_centroids <- right_join(neighb_centroids,low_ptr.25,by="neighborhood")
low_ptr_centroids$mean_lat[low_ptr_centroids$neighborhood=="Outer Mission"] <-
  low_ptr_centroids$mean_lat[low_ptr_centroids$neighborhood=="Outer Mission"] + .0045
low_ptr_centroids$mean_long[low_ptr_centroids$neighborhood=="Downtown"] <-
  low_ptr_centroids$mean_long[low_ptr_centroids$neighborhood=="Downtown"] + .0016
low_ptr_centroids$mean_lat[low_ptr_centroids$neighborhood=="Downtown"] <-
  low_ptr_centroids$mean_lat[low_ptr_centroids$neighborhood=="Downtown"] - .003

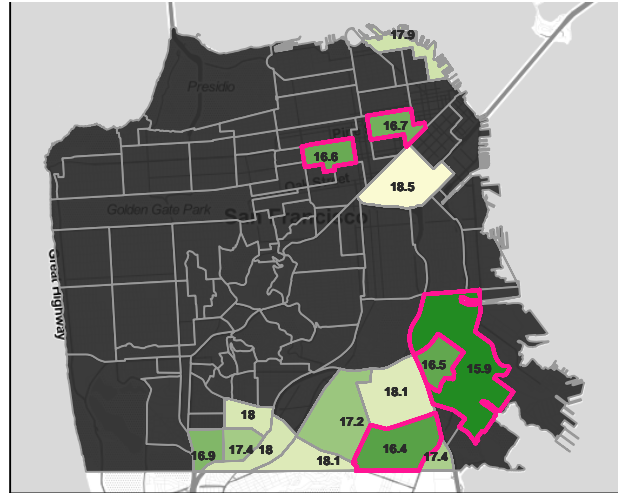
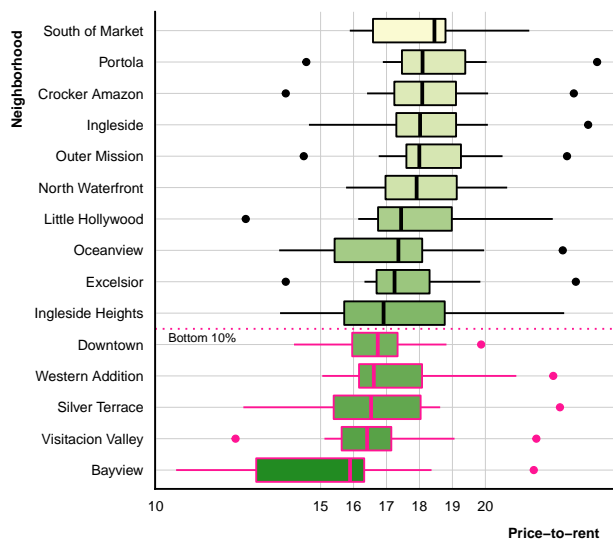
low_ptr.map <- ggmap(basemap) + coord_map() +
  geom_polygon(data=neighb_pts,
    aes(x=lat,y=long,group=neighborhood),
    color="gray60",fill="gray15",alpha=.9) +
  geom_polygon(data=filter(low_ptr_pts.25,pctile=="Max 25th percentile"),
    aes(x=lat,y=long,group=neighborhood,fill=mdn.price_to_rent),
    col="gray60") +
  geom_polygon(data=filter(low_ptr_pts.25,pctile=="Max 10th percentile"),
    aes(x=lat,y=long,group=neighborhood,fill=mdn.price_to_rent),
    col="deeppink",size=1) +
  geom_text(data=low_ptr_centroids,
    aes(x=mean_lat,y=mean_long,label=round(mdn.price_to_rent,1)),
    size=2.5,color="gray17",fontface="bold") +
  mapTheme() +
  scale_fill_gradientn(colors=c("forestgreen","lightgoldenrodyellow")) +
  theme(legend.position="none")
  #scale_color_manual(values=c("gray60","blue"))
  #theme(legend.position="none")

grid.arrange(low_ptr.boxplot,low_ptr.map,nrow=1)

```


Neighborhoods with lowest median price-to-rent

2010–2018 median yearly price-to-rent ratio (Bottom 25%)



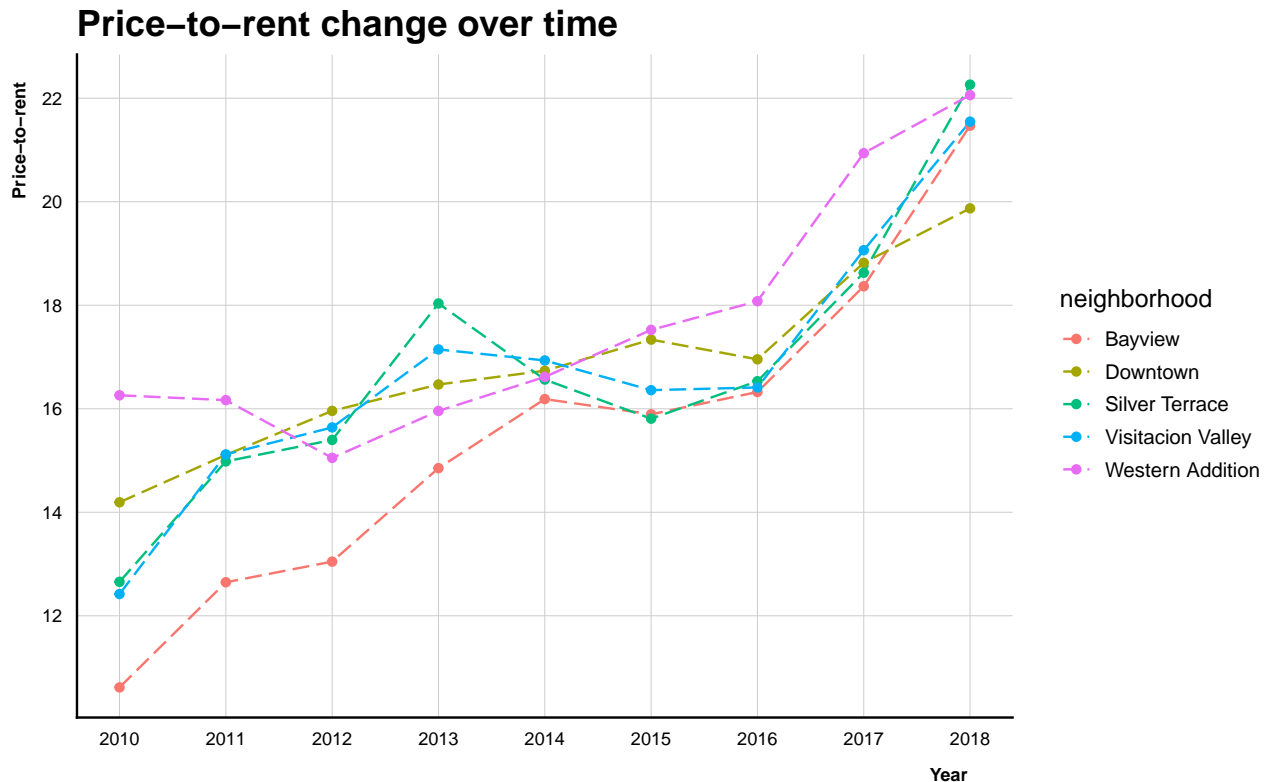
```
#ggsave("../graphics/fig7.png",width=10,height=4.9)
```

Bayview has historically high crime rates, but has grown \$400k in average home price from \$500k to \$900k based on Trulia data. There are coffee shops and eateries sprouting up in that area that cater to young professionals, making it an increasingly accessible neighborhood with a (relatively) cheap price tag.

It looks like it takes roughly under 16 years to pay off a rental purchase in that area. That's as good as it gets in the city, so consider starting your investment search there.

Next, we will look at change over time.

```
ggplot(data=low_ptr.10,aes(x=year,y=price_to_rent,color=neighborhood)) +
  geom_line(linetype="longdash") +
  geom_point() +
  plotTheme() +
  scale_x_continuous(breaks=2010:2018) +
  scale_y_continuous(breaks=seq(10,22,2)) +
  labs(x="Year",
       y="Price-to-rent",
       title="Price-to-rent change over time")
```



```
#ggsave("../graphics/fig8.png",width=8,height=5)
```

Looking at change over time, we see that Downtown actually has the lowest current price-to-rent, so we should add that to list of rental property investment candidates.

Conclusion

To recap, we've identified neighborhoods that provide high values for renters, and high returns to prospective landlords. If you're looking to rent, start your search with Saint Francis Wood, Sea Cliff, and Presidio Heights. These are affluent and safe neighborhoods and you'll be surprised to see some affordable rental candidates.

If you're looking to buy property and rent it out, you want to earn back your investment as quickly as possible. Bayview, Visitacion Valley, and Downtown offer the best return on investment for rental property estates.

This post is just one approach to thinking about housing in San Francisco. I plan to extend the housing market analysis by focusing on specific house types, examining general trends about San Francisco's housing market, and comparing housing markets in different cities.

And a note for those interested in making these graphs in R: learning how to map the neighborhoods was the trickiest part of this project. I recommend studying Ken Steif and Keith Hassel's tutorial (linked above) carefully. And learning how `geom_polygon()` works is super helpful.

If you made it this far, thanks for reading! I hope you found this interesting or helpful.

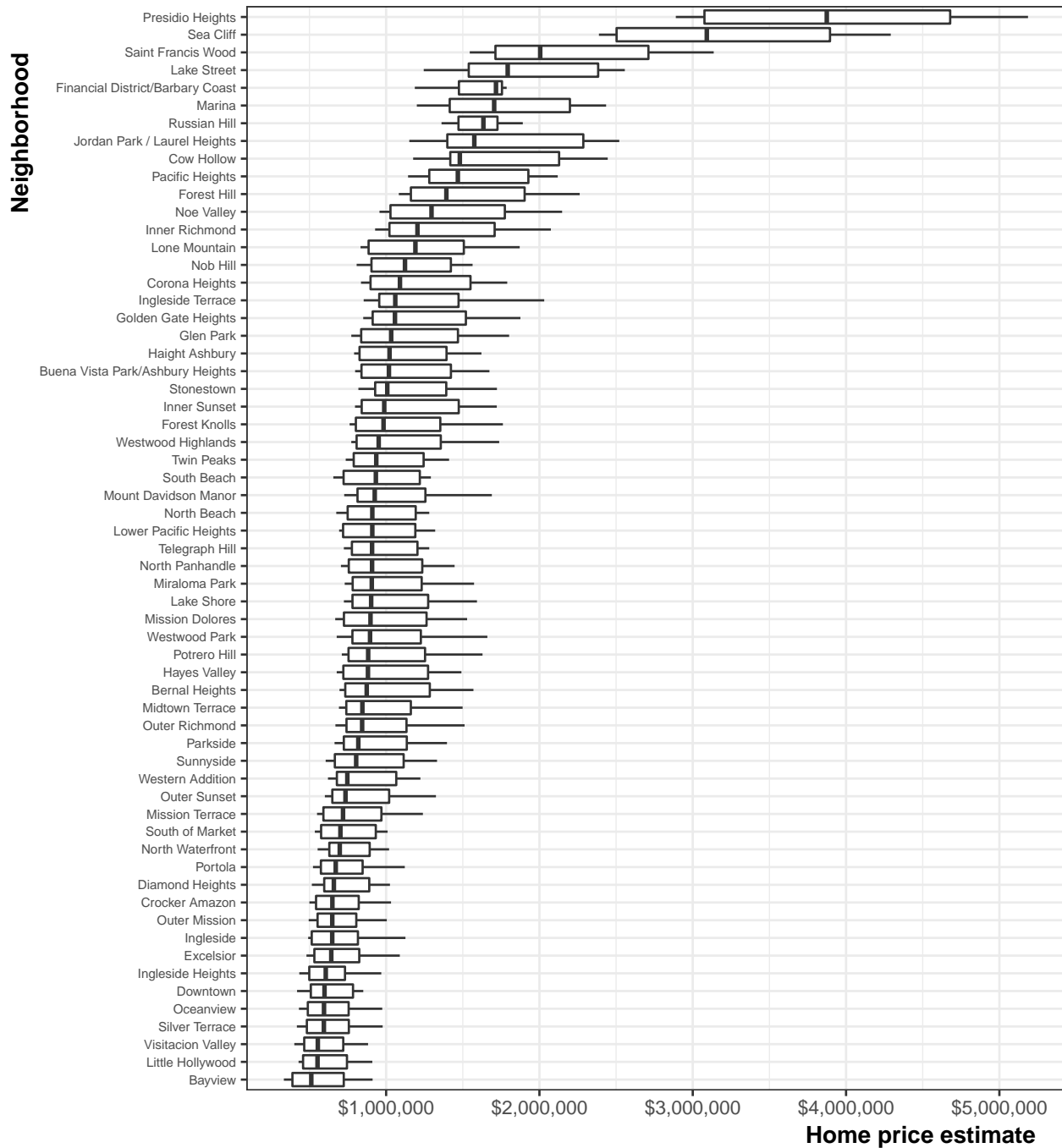
Please comment or message me feedback/constructive criticism. Always looking to improve!

Appendix

```
ggplot(filter(sf_housing,year>2007)) +  
  geom_boxplot(aes(x=reorder(neighborhood,zhvi,  
                           FUN = function(x) {median(x,na.rm=TRUE)}),  
                 y=zhvi)) +  
  labs(x="Neighborhood",y="Home price estimate",title="Home price distributions",  
       subtitle="Mean Zillow Home Value Indices >2007") +  
  scale_y_continuous(labels=dollar_format(prefix="$")) +  
  coord_flip() +  
  theme_bw() +  
  theme(axis.text.y=element_text(size=6),  
        plot.title=element_text(size=12,face="bold"),  
        plot.subtitle=element_text(size=8,face="italic"),  
        axis.title.x = element_text(hjust=.95,face="bold"),  
        axis.title.y = element_text(hjust=.95,face="bold"))
```

Home price distributions

Mean Zillow Home Value Indices >2007



```
#ggsave("../graphics/home_price.png",width=7,height=8)
```