SURV 727 Final Report

Talia Kaatz and Chris Brehm

12/13/2019

Introduction

Housing affordability and changes in the housing market garner significant attention from politicians, activists, and economists, and can be seen as indicators of broader social issues. Gentrification, predatory landlords and lenders, and segregation of housing markets as indicators of racial inequality and oppression, societal fears of economic instability manifest in predictions of the next bubble burst or housing shortages, and generational changes can be indicated by shifts in renting vs. ownership.

Of the many American metropolitan areas affected by the housing crisis, the city of Baltimore, MD provides distinct characteristics that uniquely qualify it to study as a case study. Baltimore City, unlike in many major metro areas, is not part of any surrounding counties but is classed as an "independent city." Tax bases are therefore separate at both city and county level, providing unique comparison of urban/suburban areas, as well as providing a basis for comparison between the city and other Maryland counties. Moreover, Baltimore's history of redlining and segregation within the city has produced stark demographic and economic differences at the neighborhood level, and a recent study ranked Baltimore among most heavily-gentrified cities in U.S. (Meehan 2019).

In order to explore ways of understanding the housing crisis and the factors associated with it, we have drawn data from a variety of sources including Zillow, the American Community Survey, and the Baltimore Neighborhood Indicator Alliance. Using this data, we will investigate the following hypotheses:

H1: County-level rental values will show a greater increase over time than housing values over the same period

H2: This effect will be more pronounced in urban than in rural counties

H3: Areas with younger populations will have higher rental values (compared to housing values) than areas with older populations

H4: Increases over time in home and rental values will be associated with decreases in the proportion of non-white residents

As we investigate these hypotheses, we will reflect on the challenges and advantages of using a combination of public/open access data sources to explore complex geographic/community-level data

Background

The most recent edition of the Oxford English Dictionary defines gentrification as "the process of renovating and improving housing or a district so that it conforms to middle-class taste" (Lexico 2019). Advocates and city residents often express concerns about gentrification, as it is typically associated with the displacement of poorer or non-white residents from these areas (Capps 2019, Meehan 2019). Additionally, rising rents and lack of home-ownership among millenials have led to growing concerns about rental markets, in addition to home sales markets (Hoffower 2019, Kasakove 2019).

However, gentrification is not always (perhaps not even usually) strongly associated with displacement or negative effects for minority residents (Brummet and Reed 2019). Understanding housing equity is of particular importance in the U.S., where racial segregation and housing policy have been used to restrict minority groups' access to wealth and resources (Squire and Cubrin 2005, Hoynes and McFadden 1994). Due to the complex relationships at play between time, geography, and various community demographics and characteristics, this is an area where creative data visualizations can aid in exploring and understanding patterns of housing costs, race, and age. Comparing race and age distributions in an area with average housing values or rents may each reveal a different picture of these relationships.

Data

We obtained our data from three main sources: Zillow, the American Community Survey, and the Baltimore Neighborhood Indicator Alliance. Zillow is a large online rental and real estate database company and provides tables of estimated median rent and home values by region and housing type across the U.S. These median home and rent values are calculated "in-house" monthly using Zillow's universe of housing data. These datasets are publically available and downloadable from Zillow's website. The American Community Survey (ACS) is an ongoing survey conducted by the U.S. Census Bureau. The survey gathers information on a broad range of demographic, geographic, economic, social, and housing measures. One of the datasets provided by the ACS are one-year estimates of key variables that allow analysis over a series of years. While the ACS requires an API to access, like Zillow the ACS collects data from all over the U.S. and is publically available.

Using both the Zillow and the ACS data we are able to link housing and rental prices to demographic and social variables. However, the Census does not publish 1-year estimates for areas with populations less than 65,000. So, while the 1-year estimates are useful in looking at county-level changes, it is insufficient to understanding changes in metropolitan neighborhoods. To understand how demographics in Baltimore neighborhoods have changed over time, we acquired data from Baltimore Neighborhood Indicator Alliance (BNIA). BNIA is a University of Baltimore-run group providing open-access, community-level data for Baltimore City. Freely available on the group's website is data detailing social and economic characteristics of Baltimore City's neighborhoods from 2010-2017. Additionally, the group offers geographic crosswalks and shapefiles that allow mapping of these characteristics at neighborhood level.

After obtaining these three datasets, we linked key demographic information to the Zillow dataset. The variables we pulled from the ACS were percentage white, percentage Hispanic, and median age between the years 2010-2018. We linked this data to the Zillow dataset by appropriate year and county in order to answer our research questions about the relationship between housing values and these demographic variables. At the neighborhood level, we linked BNIA data from the years 2010-2017 to neighborhoods described in the Zillow dataset. This was a careful and deliberate process as many neighborhoods were known by multiple names or encompassed other smaller neighborhoods.

In order to visualize our results with geographic maps of the regions analyzed, we linked mapping data to our complete housing and demographic data. The "maps" package in R was used for looking at Maryland at the county level while we obtained shapefiles of Baltimore City from BNIA to look at Baltimore at the neighborhood level. We then used packages such as "ggplot2" to visualize our results at multiple levels including county, metropolitan area, and neighborhood. While we were unable to formally compare different time series at this time, but it is an opportunity for future research to further explore the relationship between housing and demographic values in Maryland and Baltimore.

```
#Zillow data first, from the github repository
county_hv <- read_csv("https://raw.githubusercontent.com/taliakaatz/surv727project/master/county%20HV.cs
county_rv <- read_csv("https://raw.githubusercontent.com/taliakaatz/surv727project/master/county%20RV.cs
nhood_hv <- read_csv("https://raw.githubusercontent.com/taliakaatz/surv727project/master/neighborhood%20
nhood_rv <- read_csv("https://raw.githubusercontent.com/taliakaatz/surv727project/master/neighborhood%20

#ACS data next, using 'censusapi'
cs_key <- "a57bdfaf40991f97bfd12e41478426c81badd18a"
count_md18 <- getCensus(name = "2018/acs/acs1/profile",
                        vars = c("DP05_0037PE", "DP05_0071PE", "DP05_0018E"),#%white, %hisp, median age
                        region = "county:*",
```

```r
                      regionin = "state:24",
                      key = cs_key)[-c(1:2)]

count_md17 <- getCensus(name = "2017/acs/acs1/profile",
                      vars = c("DP05_0037PE", "DP05_0071PE", "DP05_0018E"),#%white, %hisp, median age
                      region = "county:*",
                      regionin = "state:24",
                      key = cs_key)[-c(1:2)]

count_md16 <- getCensus(name = "2016/acs/acs1/profile",
                      vars = c("DP05_0032PE", "DP05_0066PE", "DP05_0017E"),#%white, %hisp, median age
                      region = "county:*",
                      regionin = "state:24",
                      key = cs_key)[-c(1:2)]

count_md15 <- getCensus(name = "2015/acs/acs1/profile",
                      vars = c("DP05_0032PE", "DP05_0066PE", "DP05_0017E"),#%white, %hisp, median age
                      region = "county:*",
                      regionin = "state:24",
                      key = cs_key)[-c(1:2)]

count_md14 <- getCensus(name = "2014/acs/acs1/profile",
                      vars = c("DP05_0032PE", "DP05_0066PE", "DP05_0017E"),#%white, %hisp, median age
                      region = "county:*",
                      regionin = "state:24",
                      key = cs_key)[-c(1:2)]

count_md13 <- getCensus(name = "2013/acs/acs1/profile",
                      vars = c("DP05_0032PE", "DP05_0066PE", "DP05_0017E"),#%white, %hisp, median age
                      region = "county:*",
                      regionin = "state:24",
                      key = cs_key)[-c(1:2)]

count_md12 <- getCensus(name = "2012/acs/acs1/profile",
                      vars = c("DP05_0032PE", "DP05_0066PE", "DP05_0017E"),#%white, %hisp, median age
                      region = "county:*",
                      regionin = "state:24",
                      key = cs_key)[-c(1:2)]

count_md11 <- getCensus(name = "2011/acs/acs1/profile",
                      vars = c("DP05_0032PE", "DP05_0066PE", "DP05_0017E"),#%white, %hisp, median age
                      region = "county:*",
                      regionin = "state:24",
                      key = cs_key)[-c(1:2)]

count_md10 <- getCensus(name = "2010/acs/acs1/profile",
                      vars =c("DP05_0032PE", "DP05_0066PE", "DP05_0017E"),#%white, %hisp, median age
                      region = "county:*",
                      regionin = "state:24",
                      key = cs_key)[-c(1:2)]


#BNIA geographic crosswalk, also from the guthub repository
```

```r
bnia_crosswalk <- read.csv("/Users/taliakaatz/Desktop/Computation & Data Display/Project Folder/CSA_to_

#BNIA demographic data, must be downloaded from github
setwd("/Users/taliakaatz/Desktop/Computation & Data Display/Project Folder/")
bnia <- "VS17_Indicators_2010_2017.xlsx"

bnia2010 <- read_xlsx(bnia, sheet = "2010")
bnia2011 <- read_xlsx(bnia, sheet = "2011")
bnia2012 <- read_xlsx(bnia, sheet = "2012")
bnia2013 <- read_xlsx(bnia, sheet = "2013")
bnia2014 <- read_xlsx(bnia, sheet = "2014")
bnia2015 <- read_xlsx(bnia, sheet = "2015")
bnia2016 <- read_xlsx(bnia, sheet = "2016")
bnia2017 <- read_xlsx(bnia, sheet = "2017")

#BNIA shapefile for mapping
bnia_sf <- st_read("csa_2010_boundaries/CSA_NSA_Tracts.shp")

#MD county map data, from 'maps' package
md_map <- map_data("county", region = "maryland")
```

Preprocessing

County-level data

```r
RVMD <-
  county_rv %>%
    filter(State %in% "MD") %>%
    gather(key = "date", value = "RentalValue", 8:117) %>%
    select(-State, -StateCodeFIPS, -MunicipalCodeFIPS, -SizeRank) %>%
    separate(date, into = c("year", "month"), sep = "-", remove = FALSE) %>%
    filter(year %in% 2010:2018) %>%
    arrange(RegionName, desc(date))

HVMD <-
  county_hv %>%
    filter(State %in% "MD") %>%
    gather(key = "date", value = "HomeValue", 8:290) %>%
    select(-State, -StateCodeFIPS, -MunicipalCodeFIPS, -SizeRank) %>%
    separate(date, into = c("year", "month"), sep = "-", remove = FALSE) %>%
    filter(year %in% 2010:2018) %>%
    arrange(RegionName)
```

```r
count_md10$CountyName <- c("Allegany County","Anne Arundel County", "Baltimore County", "Calvert County

count_md11$CountyName <- c("Allegany County","Anne Arundel County", "Baltimore County", "Calvert County

count_md12$CountyName <- c("Allegany County","Anne Arundel County", "Baltimore County", "Calvert County

count_md13$CountyName <- c("Allegany County","Anne Arundel County", "Baltimore County", "Calvert County
```

```r
count_md14$CountyName <- c("Allegany County","Anne Arundel County", "Baltimore County", "Calvert County

count_md15$CountyName <- c("Allegany County","Anne Arundel County", "Baltimore County", "Calvert County

count_md16$CountyName <- c("Allegany County","Anne Arundel County", "Baltimore County", "Calvert County

count_md17$CountyName <- c("Allegany County","Anne Arundel County", "Baltimore County", "Calvert County

count_md18$CountyName <- c("Allegany County","Anne Arundel County", "Baltimore County", "Calvert County
```

```r
HVMD$nonwhite <- NA
HVMD$hispanic <- NA
HVMD$medianage <- NA
RVMD$nonwhite <- NA
RVMD$hispanic <- NA
RVMD$medianage <- NA

for(i in 1:length(HVMD$RegionName)){
  if(HVMD$RegionName[i] %in% count_md10$CountyName){
    if(HVMD$year[i] == 2010){
      HVMD$nonwhite[i] <- 100 - count_md10$DP05_0032PE[count_md10$CountyName == HVMD$RegionName[i]]
      HVMD$hispanic[i] <- count_md10$DP05_0066PE[count_md10$CountyName == HVMD$RegionName[i]]
      HVMD$medianage[i] <- count_md10$DP05_0017E[count_md10$CountyName == HVMD$RegionName[i]]
    }
    else if(HVMD$year[i] == 2011){
      HVMD$nonwhite[i] <- 100 - count_md11$DP05_0032PE[count_md11$CountyName == HVMD$RegionName[i]]
      HVMD$hispanic[i] <- count_md11$DP05_0066PE[count_md11$CountyName == HVMD$RegionName[i]]
      HVMD$medianage[i] <- count_md11$DP05_0017E[count_md11$CountyName == HVMD$RegionName[i]]
    }
    else if(HVMD$year[i] == 2012){
      HVMD$nonwhite[i] <- 100 - count_md12$DP05_0032PE[count_md12$CountyName == HVMD$RegionName[i]]
      HVMD$hispanic[i] <- count_md12$DP05_0066PE[count_md12$CountyName == HVMD$RegionName[i]]
      HVMD$medianage[i] <- count_md12$DP05_0017E[count_md12$CountyName == HVMD$RegionName[i]]
    }
    else if(HVMD$year[i] == 2013){
      HVMD$nonwhite[i] <- 100 - count_md13$DP05_0032PE[count_md13$CountyName == HVMD$RegionName[i]]
      HVMD$hispanic[i] <- count_md13$DP05_0066PE[count_md13$CountyName == HVMD$RegionName[i]]
      HVMD$medianage[i] <- count_md13$DP05_0017E[count_md13$CountyName == HVMD$RegionName[i]]
    }
    else if(HVMD$year[i] == 2014){
      HVMD$nonwhite[i] <- 100 - count_md14$DP05_0032PE[count_md14$CountyName == HVMD$RegionName[i]]
      HVMD$hispanic[i] <- count_md14$DP05_0066PE[count_md14$CountyName == HVMD$RegionName[i]]
      HVMD$medianage[i] <- count_md14$DP05_0017E[count_md14$CountyName == HVMD$RegionName[i]]
    }
    else if(HVMD$year[i] == 2015){
      HVMD$nonwhite[i] <- 100 - count_md15$DP05_0032PE[count_md15$CountyName == HVMD$RegionName[i]]
      HVMD$hispanic[i] <- count_md15$DP05_0066PE[count_md15$CountyName == HVMD$RegionName[i]]
      HVMD$medianage[i] <- count_md15$DP05_0017E[count_md15$CountyName == HVMD$RegionName[i]]
    }
    else if(HVMD$year[i] == 2016){
      HVMD$nonwhite[i] <- 100 - count_md16$DP05_0032PE[count_md16$CountyName == HVMD$RegionName[i]]
      HVMD$hispanic[i] <- count_md16$DP05_0066PE[count_md16$CountyName == HVMD$RegionName[i]]
      HVMD$medianage[i] <- count_md16$DP05_0017E[count_md16$CountyName == HVMD$RegionName[i]]
```

```r
    }
    else if(HVMD$year[i] == 2017){
      HVMD$nonwhite[i] <- 100 - count_md17$DP05_0037PE [count_md17$CountyName == HVMD$RegionName[i]]
      HVMD$hispanic[i] <- count_md17$DP05_0071PE[count_md17$CountyName == HVMD$RegionName[i]]
      HVMD$medianage[i] <- count_md17$DP05_0018E[count_md17$CountyName == HVMD$RegionName[i]]
    }
    else if(HVMD$year[i] == 2018){
      HVMD$nonwhite[i] <- 100 - count_md18$DP05_0037PE[count_md18$CountyName == HVMD$RegionName[i]]
      HVMD$hispanic[i] <- count_md18$DP05_0071PE[count_md18$CountyName == HVMD$RegionName[i]]
      HVMD$medianage[i] <- count_md18$DP05_0018E[count_md18$CountyName == HVMD$RegionName[i]]
    }

  }
}

for(i in 1:length(RVMD$RegionName)){
  if(RVMD$RegionName[i] %in% count_md10$CountyName){
    if(RVMD$year[i] == 2010){
      RVMD$nonwhite[i] <- 100 - count_md10$DP05_0032PE[count_md10$CountyName == RVMD$RegionName[i]]
      RVMD$hispanic[i] <- count_md10$DP05_0066PE[count_md10$CountyName == RVMD$RegionName[i]]
      RVMD$medianage[i] <- count_md10$DP05_0017E[count_md10$CountyName == RVMD$RegionName[i]]
    }
    else if(RVMD$year[i] == 2011){
      RVMD$nonwhite[i] <- 100 - count_md11$DP05_0032PE[count_md11$CountyName == RVMD$RegionName[i]]
      RVMD$hispanic[i] <- count_md11$DP05_0066PE[count_md11$CountyName == RVMD$RegionName[i]]
      RVMD$medianage[i] <- count_md11$DP05_0017E[count_md11$CountyName == RVMD$RegionName[i]]
    }
    else if(RVMD$year[i] == 2012){
      RVMD$nonwhite[i] <- 100 - count_md12$DP05_0032PE[count_md12$CountyName == RVMD$RegionName[i]]
      RVMD$hispanic[i] <- count_md12$DP05_0066PE[count_md12$CountyName == RVMD$RegionName[i]]
      RVMD$medianage[i] <- count_md12$DP05_0017E[count_md12$CountyName == RVMD$RegionName[i]]
    }
    else if(RVMD$year[i] == 2013){
      RVMD$nonwhite[i] <- 100 - count_md13$DP05_0032PE[count_md13$CountyName == RVMD$RegionName[i]]
      RVMD$hispanic[i] <- count_md13$DP05_0066PE[count_md13$CountyName == RVMD$RegionName[i]]
      RVMD$medianage[i] <- count_md13$DP05_0017E[count_md13$CountyName == RVMD$RegionName[i]]
    }
    else if(RVMD$year[i] == 2014){
      RVMD$nonwhite[i] <- 100 - count_md14$DP05_0032PE[count_md14$CountyName == RVMD$RegionName[i]]
      RVMD$hispanic[i] <- count_md14$DP05_0066PE[count_md14$CountyName == RVMD$RegionName[i]]
      RVMD$medianage[i] <- count_md14$DP05_0017E[count_md14$CountyName == RVMD$RegionName[i]]
    }
    else if(RVMD$year[i] == 2015){
      RVMD$nonwhite[i] <- 100 - count_md15$DP05_0032PE[count_md15$CountyName == RVMD$RegionName[i]]
      RVMD$hispanic[i] <- count_md15$DP05_0066PE[count_md15$CountyName == RVMD$RegionName[i]]
      RVMD$medianage[i] <- count_md15$DP05_0017E[count_md15$CountyName == RVMD$RegionName[i]]
    }
    else if(RVMD$year[i] == 2016){
      RVMD$nonwhite[i] <- 100 - count_md16$DP05_0032PE[count_md16$CountyName == RVMD$RegionName[i]]
      RVMD$hispanic[i] <- count_md16$DP05_0066PE[count_md16$CountyName == RVMD$RegionName[i]]
      RVMD$medianage[i] <- count_md16$DP05_0017E[count_md16$CountyName == RVMD$RegionName[i]]
    }
    else if(RVMD$year[i] == 2017){
```

```
      RVMD$nonwhite[i] <- 100 - count_md17$DP05_0037PE [count_md17$CountyName == RVMD$RegionName[i]]
      RVMD$hispanic[i] <- count_md17$DP05_0071PE[count_md17$CountyName == RVMD$RegionName[i]]
      RVMD$medianage[i] <- count_md17$DP05_0018E[count_md17$CountyName == RVMD$RegionName[i]]
    }
    else if(RVMD$year[i] == 2018){
      RVMD$nonwhite[i] <- 100 - count_md18$DP05_0037PE[count_md18$CountyName == RVMD$RegionName[i]]
      RVMD$hispanic[i] <- count_md18$DP05_0071PE[count_md18$CountyName == RVMD$RegionName[i]]
      RVMD$medianage[i] <- count_md18$DP05_0018E[count_md18$CountyName == RVMD$RegionName[i]]
    }

  }
}
HVMD$nonwhite[HVMD$nonwhite > 100] <- NA
RVMD$nonwhite[RVMD$nonwhite > 100] <- NA

combinedMD <-
  HVMD %>%
    left_join(RVMD, by = c("RegionName", "date"))
```

Preprocessing

Neighborhood-level data

```
RVBM  <-
  nhood_rv %>%
    filter(City %in% "Baltimore" & State %in% "MD") %>%
    gather(key = "date", value = "RentalValue", 8:117) %>%
    select(-State, -Metro, -City, -CountyName) %>%
    separate(date, into = c("year", "month"), sep = "-", remove = FALSE) %>%
    filter(year %in% 2010:2018) %>%
    arrange(RegionName, desc(date))

HVBM <-
  nhood_hv %>%
    filter(City %in% "Baltimore") %>%
    gather(key = "date", value = "HomeValue", 8:290) %>%
    select(-State, -Metro, -City, -CountyName) %>%
    separate(date, into = c("year", "month"), sep = "-", remove = FALSE) %>%
    filter(year %in% 2010:2018) %>%
    arrange(RegionName, desc(date))

#check whether neighborhood names line up between rental and home value files

countynames <- unique(RVBM$RegionName)
rent_hoods <- countynames[!(countynames %in% HVBM$RegionName)]

countynames2 <- unique(HVBM$RegionName)
home_hoods <- countynames2[!(countynames2 %in% RVBM$RegionName)]

print(rent_hoods)
```

```r
print(home_hoods)
#evidence of significant differences in the neighborhoods present in each file
```

Standardize crosswalk and zillow files and merge by neighborhood

```r
#create uniform neighborhood name variables in each dataset

bnia_crosswalk %<>%
  mutate(NHOOD = trimws(toupper(NSA2010))) %>%
  select(-NSA2010)

HVBM %<>%
  mutate(NHOOD = trimws(toupper(RegionName)))

RVBM %<>%
  mutate(NHOOD = trimws(toupper(RegionName)))

#merge crosswalk to home and rental value data and identify unmatched neighborhood names
HVBM %>%
  left_join(bnia_crosswalk, by = "NHOOD") %>%
  filter(is.na(CSA2010)) %>%
  distinct(NHOOD)

RVBM %>%
  left_join(bnia_crosswalk, by = "NHOOD") %>%
  filter(is.na(CSA2010)) %>%
  distinct(NHOOD)

#fix unmatched neighborhood names (update in Zillow file to match crosswalk):

HVBM %<>%
  mutate(NHOOD = gsub("CHINQUAPIN PARK-BELVEDERE", "CHINQUAPIN PARK", NHOOD)) %>%
  mutate(NHOOD = gsub("HARFORD-ECHODALE - PERRING PARKWAY", "HARFORD-ECHODALE/PERRING PARKWAY", NHOOD))
  mutate(NHOOD = gsub("JOSEPH LEE", "BAYVIEW", NHOOD)) %>%
  mutate(NHOOD = gsub("MOUNT WASHINGTON", "MT. WASHINGTON", NHOOD)) %>%
  mutate(NHOOD = gsub("NORTH ROLAND PARK - POPLAR HILL", "NORTH ROLAND PARK/POPLAR HILL", NHOOD)) %>%
  mutate(NHOOD = gsub("SBIC", "SOUTH BALTIMORE", NHOOD)) %>%
  mutate(NHOOD = gsub("WOODRING", "WESTFIELD", NHOOD))


RVBM %<>%
  mutate(NHOOD = gsub("COPPIN HEIGHTS - ASH-CO-EAST", "COPPIN HEIGHTS/ASH-CO-EAST", NHOOD)) %>%
  mutate(NHOOD = gsub("ELLWOOD PARK-MONUMENT", "ELLWOOD PARK/MONUMENT", NHOOD)) %>%
  mutate(NHOOD = gsub("HARFORD-ECHODALE - PERRING PARKWAY", "HARFORD-ECHODALE/PERRING PARKWAY", NHOOD))
  mutate(NHOOD = gsub("MONDAWIN", "MONDAWMIN", NHOOD)) %>%
  mutate(NHOOD = gsub("MOUNT WASHINGTON", "MT. WASHINGTON", NHOOD)) %>%
  mutate(NHOOD = gsub("NEW SOUTHWEST - MOUNT CLARE", "NEW SOUTHWEST/MT. CLARE", NHOOD)) %>%
  mutate(NHOOD = gsub("PENROSE-FAYETTE STREET OUTREACH", "PENROSE/FAYETTE STREET OUTREACH", NHOOD)) %>%
  mutate(NHOOD = gsub("SBIC", "SOUTH BALTIMORE", NHOOD)) %>%
  mutate(NHOOD = gsub("WASHINGTON VILLAGE", "WASHINGTON VILLAGE/PIGTOWN", NHOOD)) %>%
  mutate(NHOOD = gsub("WOODRING", "WESTFIELD", NHOOD))

#merge and check again with updated neighborhoods -- looks good!
```

```r
HVBM %>%
  left_join(bnia_crosswalk, by = "NHOOD") %>%
  filter(is.na(CSA2010)) %>%
  distinct(NHOOD)

RVBM %>%
  left_join(bnia_crosswalk, by = "NHOOD") %>%
  filter(is.na(CSA2010)) %>%
  distinct(NHOOD)

#merge CSA names from crosswalk to datasets

HVBM %<>% left_join(bnia_crosswalk, by = "NHOOD")

RVBM %<>% left_join(bnia_crosswalk, by = "NHOOD")

#adjust some CSA names to match formatting to BNIA demographic files

RVBM %<>%
  mutate(CSA2010 = gsub("Allendale/Irvington/South Hilton", "Allendale/Irvington/S. Hilton", CSA2010)) %
  mutate(CSA2010 = gsub("Glen-Falstaff", "Glen-Fallstaff", CSA2010)) %>%
  mutate(CSA2010 = gsub("Westport/Mt. Winans/Lakeland", "Westport/Mount Winans/Lakeland", CSA2010)) %>%
  mutate(CSA2010 = gsub("Mt. Washington/Coldspring", "Mount Washington/Coldspring", CSA2010))

HVBM %<>%
  mutate(CSA2010 = gsub("Allendale/Irvington/South Hilton", "Allendale/Irvington/S. Hilton", CSA2010)) %
  mutate(CSA2010 = gsub("Glen-Falstaff", "Glen-Fallstaff", CSA2010)) %>%
  mutate(CSA2010 = gsub("Westport/Mt. Winans/Lakeland", "Westport/Mount Winans/Lakeland", CSA2010)) %>%
  mutate(CSA2010 = gsub("Mt. Washington/Coldspring", "Mount Washington/Coldspring", CSA2010))
```

Prepare BNIA demographic files for merge

```r
#in file for each year, subset to variables of interest and rename

#2010
bnia_vars <- colnames(bnia2010[, c("CSA2010", "Percent of Residents - White/Caucasian (Non-Hispanic)",
new_varnames <- c("CSA2010", "nonwhite10", "hispanic10", "percentya10", "mediansold10", "ai_mortgage10"

bnia2010 %<>%
  select(bnia_vars) %>%
  set_colnames(new_varnames) %>%
  mutate(nonwhite10 = (100 - nonwhite10))

#2011
bnia_vars <- c("CSA2010", "Median Price of Homes Sold", "Affordability Index - Mortgage", "Affordabilit
new_varnames <- c("CSA2010", "mediansold11", "ai_mortgage11", "ai_rent11")

bnia2011 %<>%
  select(bnia_vars) %>%
  set_colnames(new_varnames)

#2012
bnia_vars <- c("CSA2010", "Median Price of Homes Sold", "Affordability Index - Mortgage", "Affordabilit
```

```r
new_varnames <- c("CSA2010", "mediansold12", "ai_mortgage12", "ai_rent12")

bnia2012 %<>%
  select(bnia_vars) %>%
  set_colnames(new_varnames)

#2013
bnia_vars <- c("CSA2010", "Median Price of Homes Sold", "Affordability Index - Mortgage", "Affordabilit
new_varnames <- c("CSA2010", "mediansold13", "ai_mortgage13", "ai_rent13")

bnia2013 %<>%
  select(bnia_vars) %>%
  set_colnames(new_varnames)

#2014
bnia_vars <- c("CSA2010", "Median Price of Homes Sold", "Affordability Index - Mortgage", "Affordabilit
new_varnames <- c("CSA2010", "mediansold14", "ai_mortgage14", "ai_rent14")

bnia2014 %<>%
  select(bnia_vars) %>%
  set_colnames(new_varnames)

#2015
bnia_vars <- c("CSA2010", "Median Price of Homes Sold", "Affordability Index - Mortgage", "Affordabilit
new_varnames <- c("CSA2010", "mediansold15", "ai_mortgage15", "ai_rent15")

bnia2015 %<>%
  select(bnia_vars) %>%
  set_colnames(new_varnames)

#2016
bnia_vars <- colnames(bnia2016[, c("CSA2010", "Percent of Residents - White/Caucasian (Non-Hispanic)",
new_varnames <- c("CSA2010", "nonwhite16", "hispanic16", "percentya16", "mediansold16", "ai_mortgage16"

bnia2016 %<>%
  select(bnia_vars) %>%
  set_colnames(new_varnames) %>%
  mutate(nonwhite16 = (100 - nonwhite16))

#2017
bnia_vars <- colnames(bnia2017[, c("CSA2010", "Percent of Residents - White/Caucasian (Non-Hispanic)",
new_varnames <- c("CSA2010", "nonwhite17", "hispanic17", "percentya17", "mediansold17", "ai_mortgage17"

bnia2017 %<>%
  select(bnia_vars) %>%
  set_colnames(new_varnames) %>%
  mutate(nonwhite17 = (100 - nonwhite17))
```

```r
#merge BNIA to Zillow files

RVBM %<>%
  left_join(bnia2010, by = "CSA2010") %>%
  left_join(bnia2011, by = "CSA2010") %>%
```

```r
  left_join(bnia2012, by = "CSA2010") %>%
  left_join(bnia2013, by = "CSA2010") %>%
  left_join(bnia2014, by = "CSA2010") %>%
  left_join(bnia2015, by = "CSA2010") %>%
  left_join(bnia2016, by = "CSA2010") %>%
  left_join(bnia2017, by = "CSA2010")

HVBM %<>%
  left_join(bnia2010, by = "CSA2010") %>%
  left_join(bnia2011, by = "CSA2010") %>%
  left_join(bnia2012, by = "CSA2010") %>%
  left_join(bnia2013, by = "CSA2010") %>%
  left_join(bnia2014, by = "CSA2010") %>%
  left_join(bnia2015, by = "CSA2010") %>%
  left_join(bnia2016, by = "CSA2010") %>%
  left_join(bnia2017, by = "CSA2010")
```

```r
#prepare "long-format" demographic variables

RVBM$nonwhite <- ifelse(RVBM$year == 2010, RVBM$nonwhite10,
                        ifelse(RVBM$year == 2016, RVBM$nonwhite16,
                               ifelse(RVBM$year == 2017, RVBM$nonwhite17, "NA")))

RVBM$hispanic <- ifelse(RVBM$year == 2010, RVBM$hispanic10,
                        ifelse(RVBM$year == 2016, RVBM$hispanic16,
                               ifelse(RVBM$year == 2017, RVBM$hispanic17, "NA")))

RVBM$percentya <- ifelse(RVBM$year == 2010, RVBM$percentya10,
                        ifelse(RVBM$year == 2016, RVBM$percentya16,
                               ifelse(RVBM$year == 2017, RVBM$percentya17, "NA")))

RVBM$mediansold <- ifelse(RVBM$year == 2010, RVBM$mediansold10,
                          ifelse(RVBM$year == 2011, RVBM$mediansold11,
                            ifelse(RVBM$year == 2012, RVBM$mediansold12,
                              ifelse(RVBM$year == 2013, RVBM$mediansold13,
                                ifelse(RVBM$year == 2014, RVBM$mediansold14,
                                  ifelse(RVBM$year == 2015, RVBM$mediansold15,
                                   ifelse(RVBM$year == 2016, RVBM$mediansold16,
                                      ifelse(RVBM$year == 2017, RVBM$mediansold17, "NA"))))))))

RVBM$ai_mortgage <- ifelse(RVBM$year == 2010, RVBM$ai_mortgage10,
                          ifelse(RVBM$year == 2011, RVBM$ai_mortgage11,
                            ifelse(RVBM$year == 2012, RVBM$ai_mortgage12,
                              ifelse(RVBM$year == 2013, RVBM$ai_mortgage13,
                                ifelse(RVBM$year == 2014, RVBM$ai_mortgage14,
                                  ifelse(RVBM$year == 2015, RVBM$ai_mortgage15,
                                   ifelse(RVBM$year == 2016, RVBM$ai_mortgage16,
                                      ifelse(RVBM$year == 2017, RVBM$ai_mortgage17, "NA"))))))))

RVBM$ai_rent <- ifelse(RVBM$year == 2010, RVBM$ai_rent10,
                          ifelse(RVBM$year == 2011, RVBM$ai_rent11,
                            ifelse(RVBM$year == 2012, RVBM$ai_rent12,
                              ifelse(RVBM$year == 2013, RVBM$ai_rent13,
```

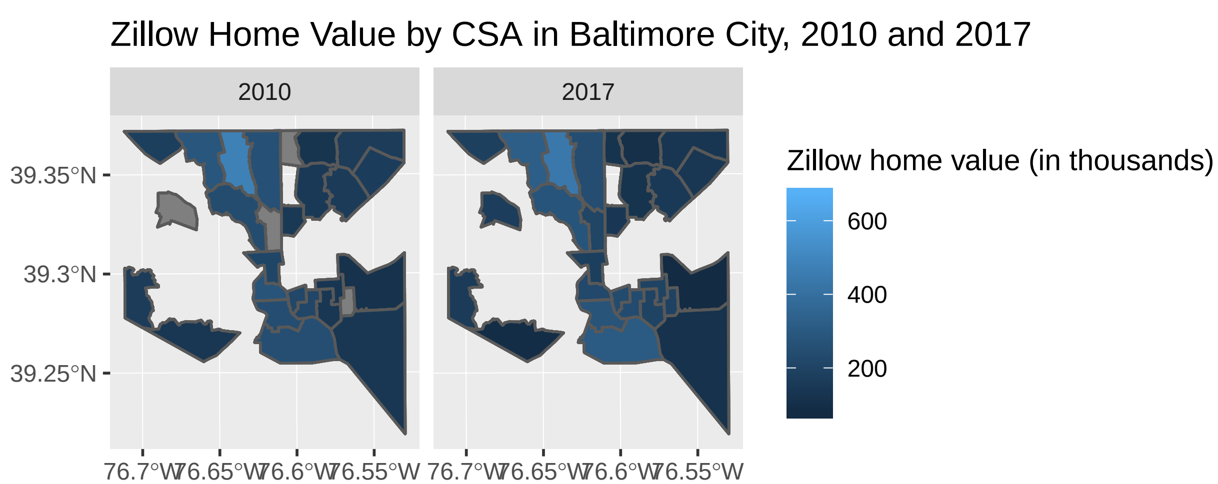
```
                                ifelse(RVBM$year == 2014, RVBM$ai_rent14,
                                  ifelse(RVBM$year == 2015, RVBM$ai_rent15,
                                   ifelse(RVBM$year == 2016, RVBM$ai_rent16,
                                      ifelse(RVBM$year == 2017, RVBM$ai_rent17, "NA"))))))))

#HVBM
HVBM$nonwhite <- ifelse(HVBM$year == 2010, HVBM$nonwhite10,
                        ifelse(HVBM$year == 2016, HVBM$nonwhite16,
                               ifelse(HVBM$year == 2017, HVBM$nonwhite17, "NA")))

HVBM$hispanic <- ifelse(HVBM$year == 2010, HVBM$hispanic10,
                        ifelse(HVBM$year == 2016, HVBM$hispanic16,
                               ifelse(HVBM$year == 2017, HVBM$hispanic17, "NA")))

HVBM$percentya <- ifelse(HVBM$year == 2010, HVBM$percentya10,
                        ifelse(HVBM$year == 2016, HVBM$percentya16,
                               ifelse(HVBM$year == 2017, HVBM$percentya17, "NA")))

HVBM$mediansold <- ifelse(HVBM$year == 2010, HVBM$mediansold10,
                          ifelse(HVBM$year == 2011, HVBM$mediansold11,
                            ifelse(HVBM$year == 2012, HVBM$mediansold12,
                              ifelse(HVBM$year == 2013, HVBM$mediansold13,
                                ifelse(HVBM$year == 2014, HVBM$mediansold14,
                                  ifelse(HVBM$year == 2015, HVBM$mediansold15,
                                   ifelse(HVBM$year == 2016, HVBM$mediansold16,
                                      ifelse(HVBM$year == 2017, HVBM$mediansold17, "NA"))))))))

HVBM$ai_mortgage <- ifelse(HVBM$year == 2010, HVBM$ai_mortgage10,
                          ifelse(HVBM$year == 2011, HVBM$ai_mortgage11,
                            ifelse(HVBM$year == 2012, HVBM$ai_mortgage12,
                              ifelse(HVBM$year == 2013, HVBM$ai_mortgage13,
                                ifelse(HVBM$year == 2014, HVBM$ai_mortgage14,
                                  ifelse(HVBM$year == 2015, HVBM$ai_mortgage15,
                                   ifelse(HVBM$year == 2016, HVBM$ai_mortgage16,
                                      ifelse(HVBM$year == 2017, HVBM$ai_mortgage17, "NA"))))))))

HVBM$ai_rent <- ifelse(HVBM$year == 2010, HVBM$ai_rent10,
                          ifelse(HVBM$year == 2011, HVBM$ai_rent11,
                            ifelse(HVBM$year == 2012, HVBM$ai_rent12,
                              ifelse(HVBM$year == 2013, HVBM$ai_rent13,
                                ifelse(HVBM$year == 2014, HVBM$ai_rent14,
                                  ifelse(HVBM$year == 2015, HVBM$ai_rent15,
                                   ifelse(HVBM$year == 2016, HVBM$ai_rent16,
                                      ifelse(HVBM$year == 2017, HVBM$ai_rent17, "NA"))))))))

RVBM %<>%
  mutate_at(c("nonwhite", "hispanic", "percentya", "mediansold", "ai_mortgage", "ai_rent"), as.numeric)

HVBM %<>%
  mutate_at(c("nonwhite", "hispanic", "percentya", "mediansold", "ai_mortgage", "ai_rent"), as.numeric)
```

Results

Variable Distribution Maps

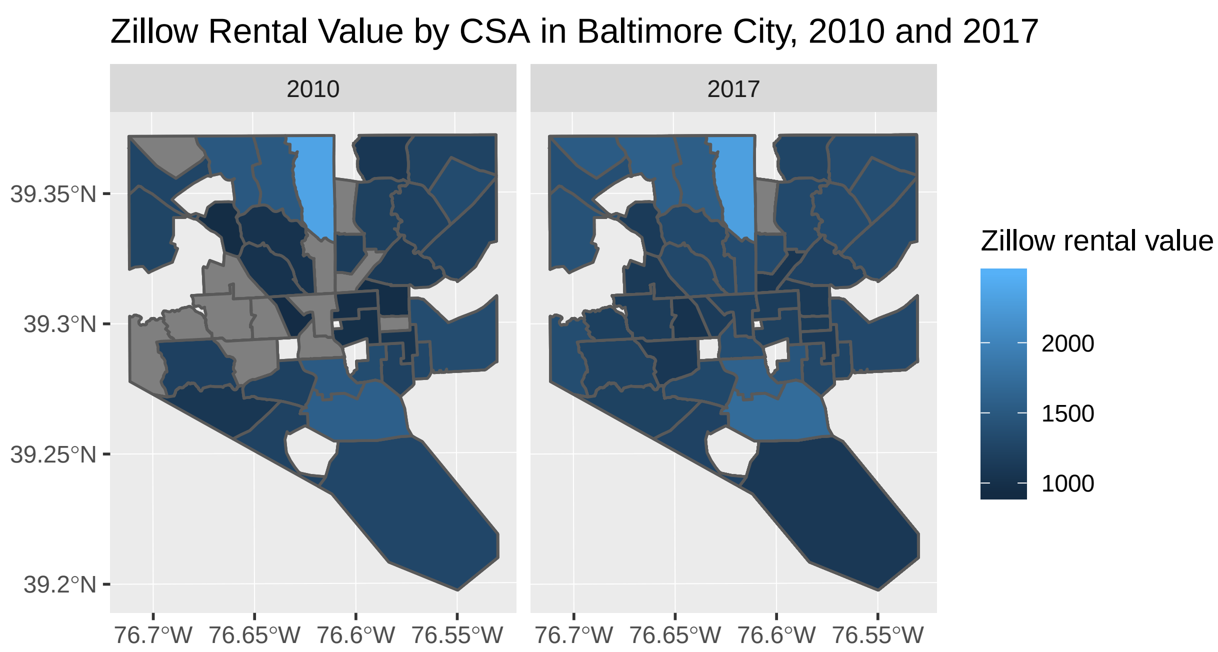
The first two maps, of the Zillow home and rent values, primarily indicate where Zillow has missing data – the east and west areas of the city are sparsely populated, particularly the home value data. The following maps illustrate how this pattern of missingness tends to align with home and rental values, the affordability indices, and the nonwhite proportion of the population. Note that Zillow data is aggregated at the year and CSA level, which may lead to a kind of "bias" in the mapping if housing/rental values are very unevenly distributed across neighborhoods within a given CSA.

```
bnia_sf %>%
  mutate(CSA2010 = Community) %>%
  left_join(HVBM, by = "CSA2010") %>%
  filter(year == c("2017", "2010")) %>%
  ggplot() +
    geom_sf(aes(fill = (HomeValue/1000))) +
    ggtitle("Zillow Home Value by CSA in Baltimore City, 2010 and 2017") +
    scale_fill_continuous(name = "Zillow home value (in thousands)") +
    facet_wrap(~year)
```

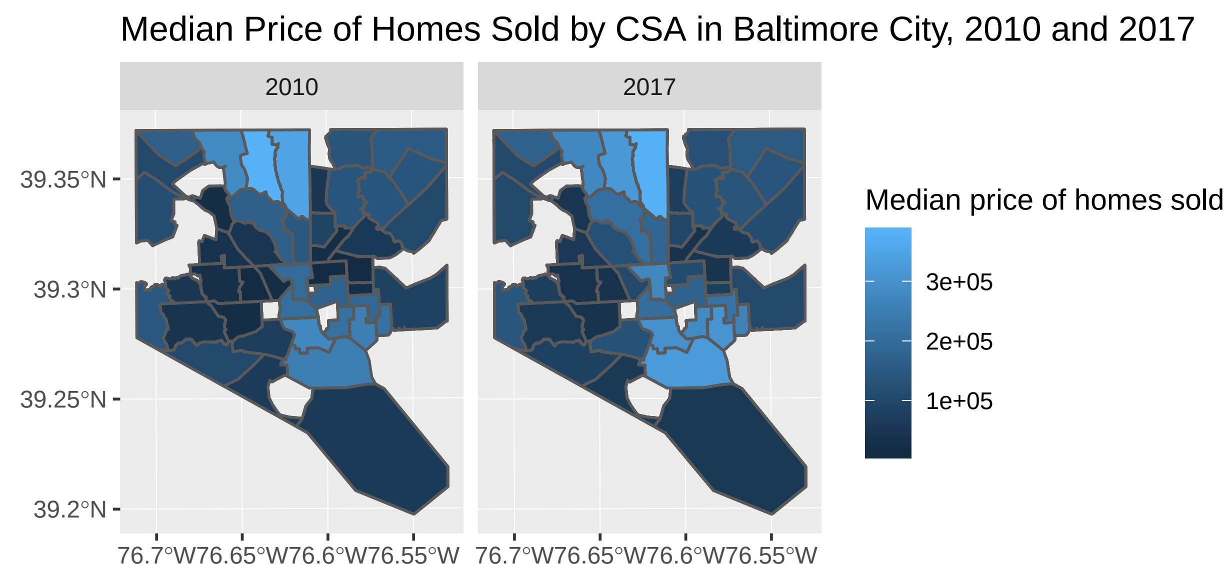



```
bnia_sf %>%
  mutate(CSA2010 = Community) %>%
  left_join(RVBM, by = "CSA2010") %>%
  filter(year == c("2017", "2010")) %>%
```

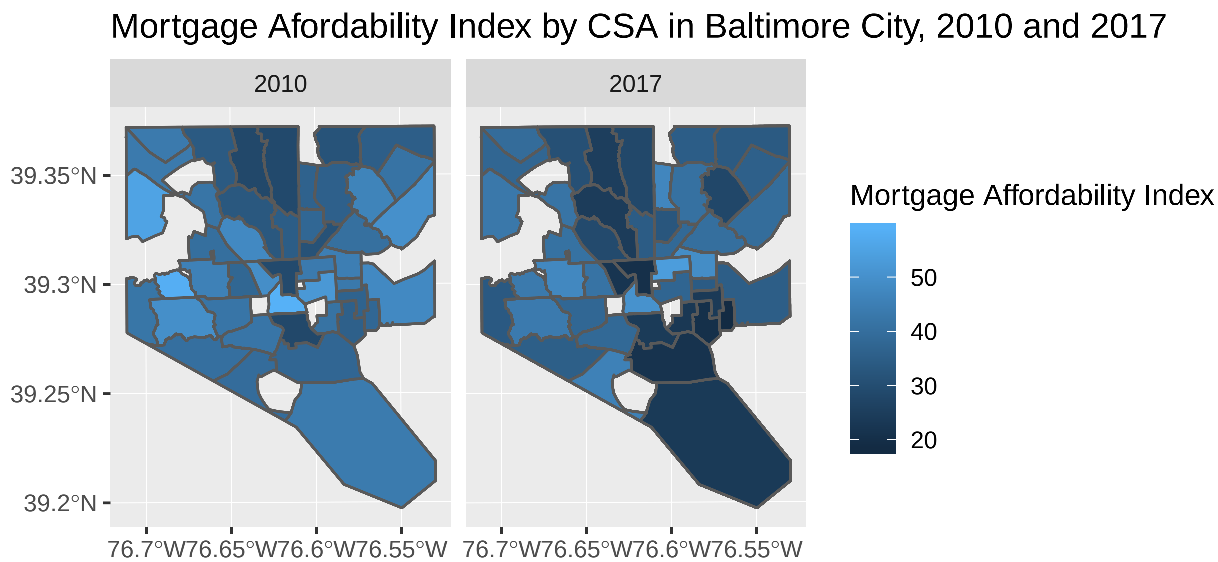
```r
  ggplot() +
    geom_sf(aes(fill = RentalValue)) +
    ggtitle("Zillow Rental Value by CSA in Baltimore City, 2010 and 2017") +
    scale_fill_continuous(name = "Zillow rental value") +
    facet_wrap(~year)
```



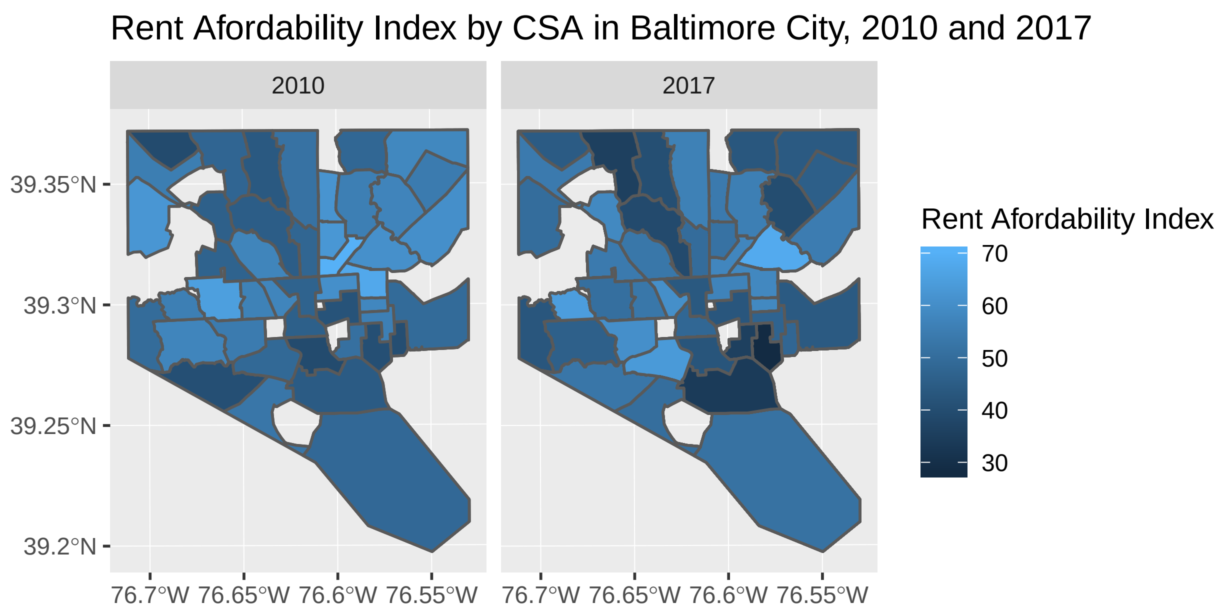

```r
bnia_sf %>%
  mutate(CSA2010 = Community) %>%
  left_join(RVBM, by = "CSA2010") %>%
  filter(year == c("2017", "2010")) %>%
  ggplot() +
    geom_sf(aes(fill = mediansold)) +
    ggtitle("Median Price of Homes Sold by CSA in Baltimore City, 2010 and 2017") +
    scale_fill_continuous(name = "Median price of homes sold") +
    facet_wrap(~year)
```



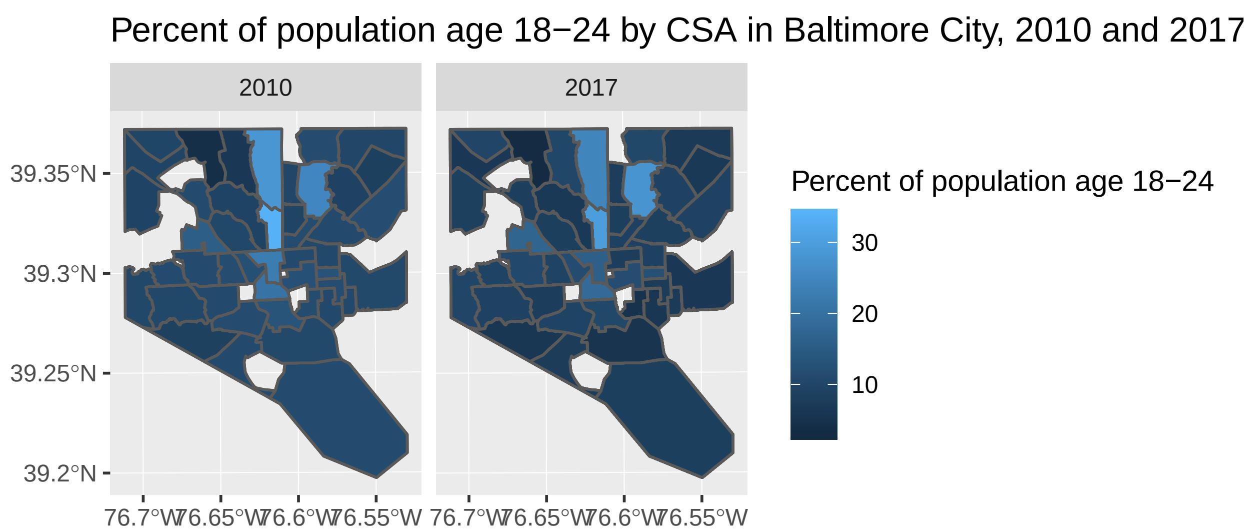

```r
bnia_sf %>%
  mutate(CSA2010 = Community) %>%
  left_join(RVBM, by = "CSA2010") %>%
  filter(year == c("2017", "2010")) %>%
  ggplot() +
    geom_sf(aes(fill = ai_mortgage)) +
    ggtitle("Mortgage Afordability Index by CSA in Baltimore City, 2010 and 2017") +
    scale_fill_continuous(name = "Mortgage Affordability Index") +
    facet_wrap(~year)
```



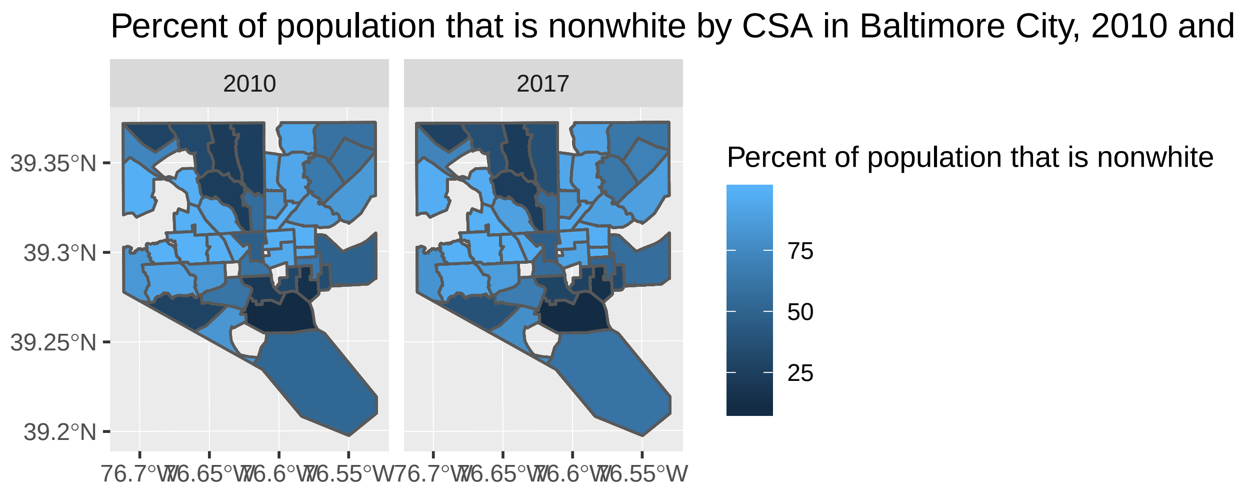

```r
bnia_sf %>%
  mutate(CSA2010 = Community) %>%
  left_join(RVBM, by = "CSA2010") %>%
  filter(year == c("2017", "2010")) %>%
  ggplot() +
    geom_sf(aes(fill = ai_rent)) +
    ggtitle("Rent Afordability Index by CSA in Baltimore City, 2010 and 2017") +
    scale_fill_continuous(name = "Rent Afordability Index") +
    facet_wrap(~year)
```




```r
bnia_sf %>%
  mutate(CSA2010 = Community) %>%
  left_join(RVBM, by = "CSA2010") %>%
  filter(year == c("2017", "2010")) %>%
  ggplot() +
    geom_sf(aes(fill = percentya)) +
    ggtitle("Percent of population age 18-24 by CSA in Baltimore City, 2010 and 2017") +
    scale_fill_continuous(name = "Percent of population age 18-24") +
    facet_wrap(~year)
```




```r
bnia_sf %>%
  mutate(CSA2010 = Community) %>%
  left_join(RVBM, by = "CSA2010") %>%
  filter(year == c("2017", "2010")) %>%
  ggplot() +
    geom_sf(aes(fill = nonwhite)) +
    ggtitle("Percent of population that is nonwhite by CSA in Baltimore City, 2010 and 2017") +
    scale_fill_continuous(name = "Percent of population that is nonwhite") +
    facet_wrap(~year)
```



Results

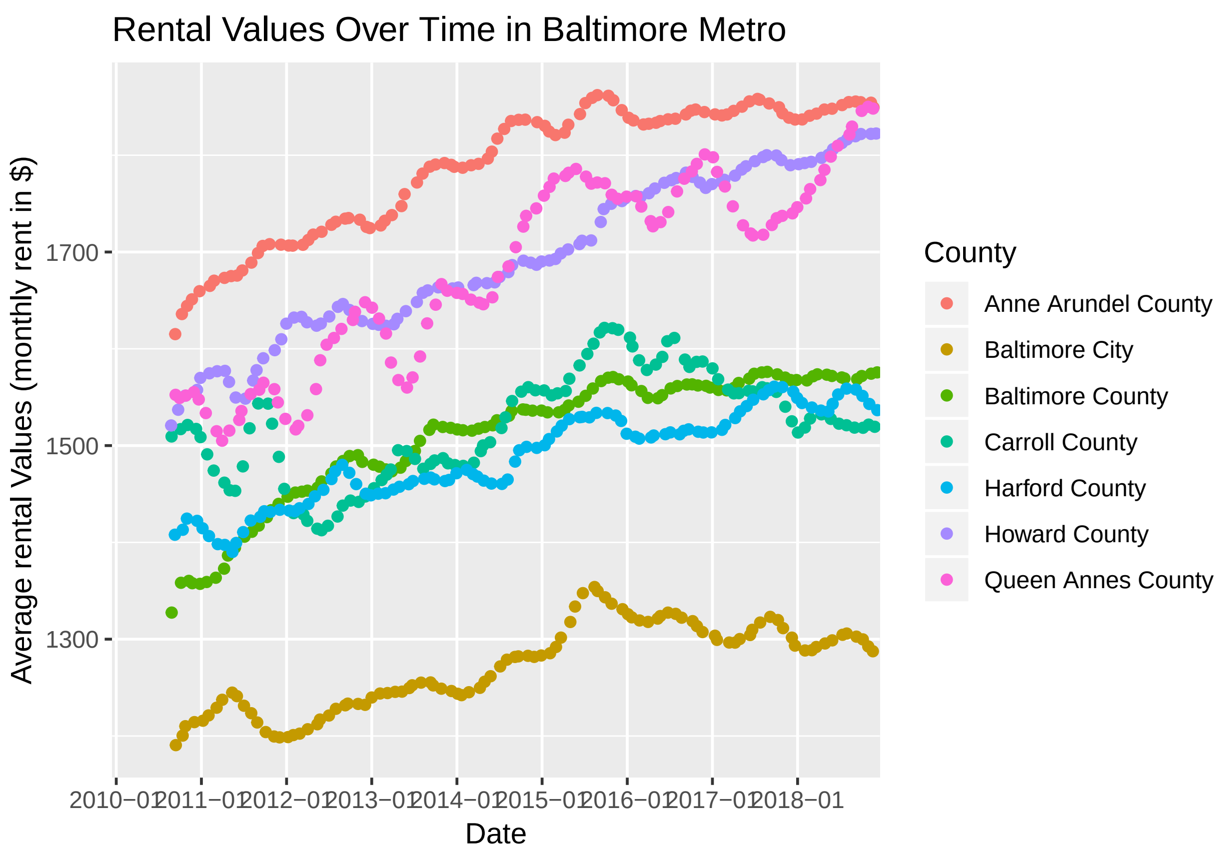
H1: County-level rental values will show a greater increase over time than housing values over the same period

Baltimore City, shaded yellow, has distinctly lower home and rental values over time compared to the suburban and rural counties surrounding it. Interestingly, the county with the highest rents (Anne Arundel) did not have the highest home values (Howard). Notably, the home and rental values show similar patterns over time across counties, and Baltimore City looks similar to the other counties.

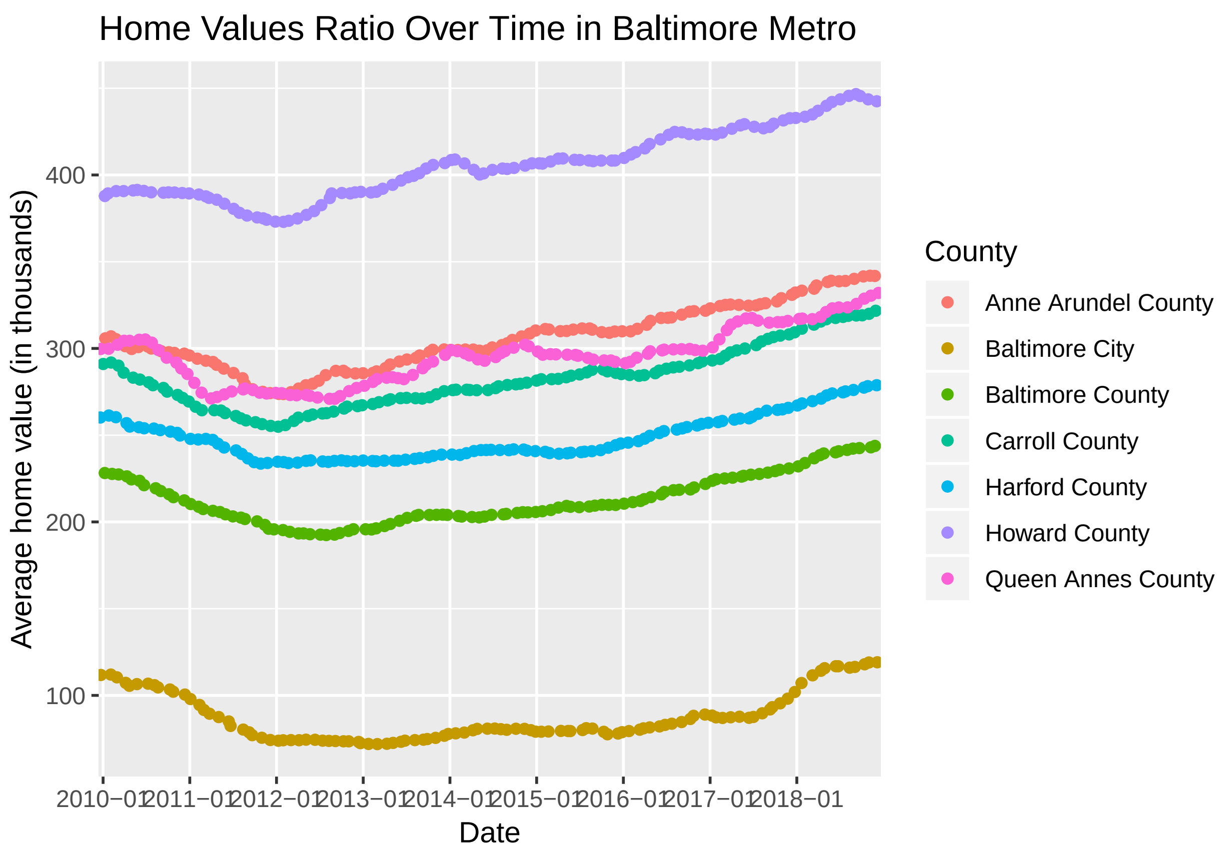
However, when rent and home values are visualized as a ratio of (Rental Value/Home Value in thousands), Baltimore City shows very different change over time compared to other counties, with rent getting more expensive relative to home values until steadying around 2013 and then dropping starting in 2016. The ratio appears to be significantly lower in the other counties, where it remains nearly flat between 2011 and 2019. These results do not provide evidence in support of Hypothesis 1, though they do show that rents are generally higher relative to home values in urban Baltimore than the suburban/rural counties.

```r
year_breaks <- c("2010-01", "2011-01", "2012-01", "2013-01", "2014-01", "2015-01", "2016-01", "2017-01"

combinedMD %>%
  filter(Metro.x %in% "Baltimore-Columbia-Towson") %>%
  ggplot() +
  geom_jitter(aes(x = date, y = RentalValue , group = RegionName, color = RegionName)) +
  ggtitle("Rental Values Over Time in Baltimore Metro") +
```

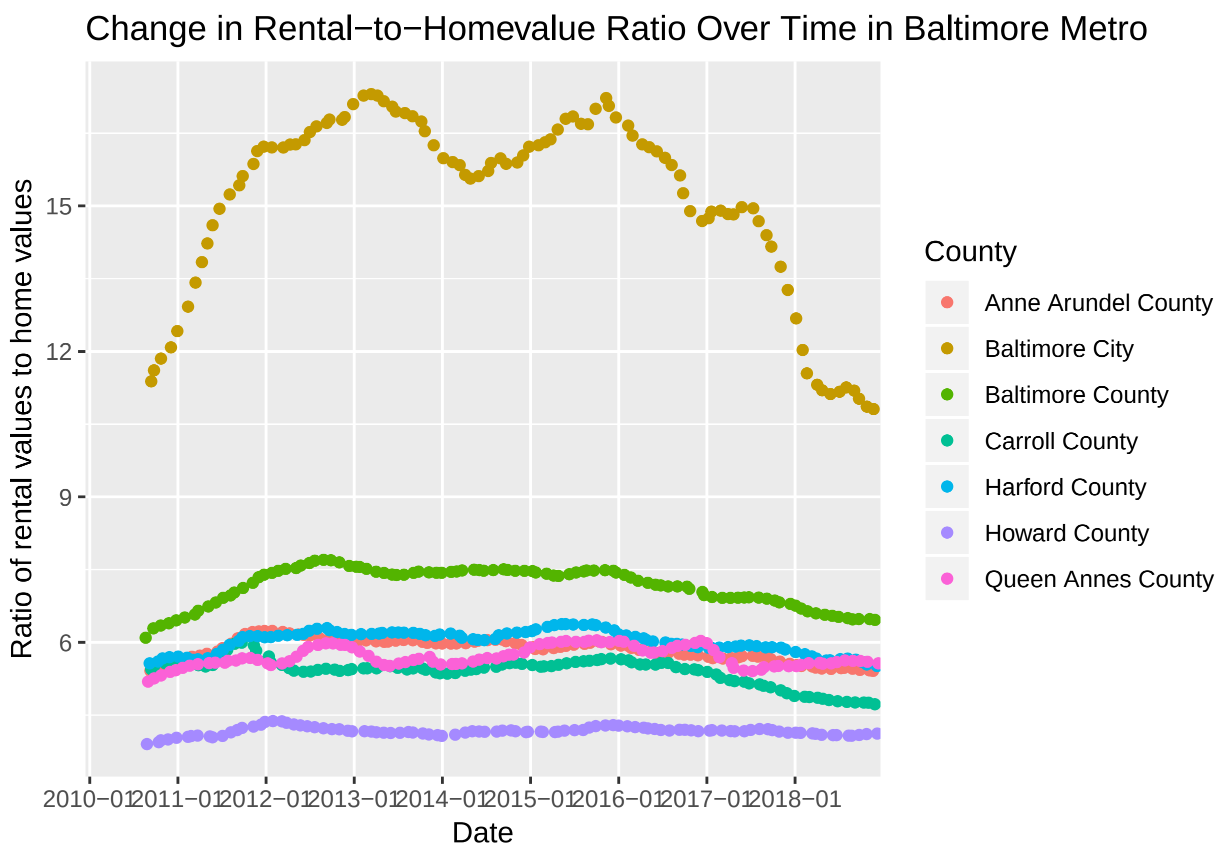
```
  scale_x_discrete("Date", c('2010', year_breaks, year_breaks)) +
  ylab("Average rental Values (monthly rent in $)") +
  scale_color_discrete("County")
```




```
combinedMD %>%
  filter(Metro.x %in% "Baltimore-Columbia-Towson") %>%
  ggplot() +
  geom_jitter(aes(x = date, y = (HomeValue/1000) , group = RegionName, color = RegionName)) +
  ggtitle("Home Values Ratio Over Time in Baltimore Metro") +
  scale_x_discrete("Date", c('2010', year_breaks, year_breaks)) +
  ylab("Average home value (in thousands)") +
  scale_color_discrete("County")
```




```
combinedMD %>%
  mutate(hrv_ratio = (RentalValue/(HomeValue/1000))) %>%
  filter(Metro.x %in% "Baltimore-Columbia-Towson") %>%
  ggplot() +
  geom_jitter(aes(x = date, y = hrv_ratio , group = RegionName, color = RegionName)) +
  ggtitle("Change in Rental-to-Homevalue Ratio Over Time in Baltimore Metro") +
  scale_x_discrete("Date", c('2010', year_breaks, year_breaks)) +
  ylab("Ratio of rental values to home values") +
  scale_color_discrete("County")
```

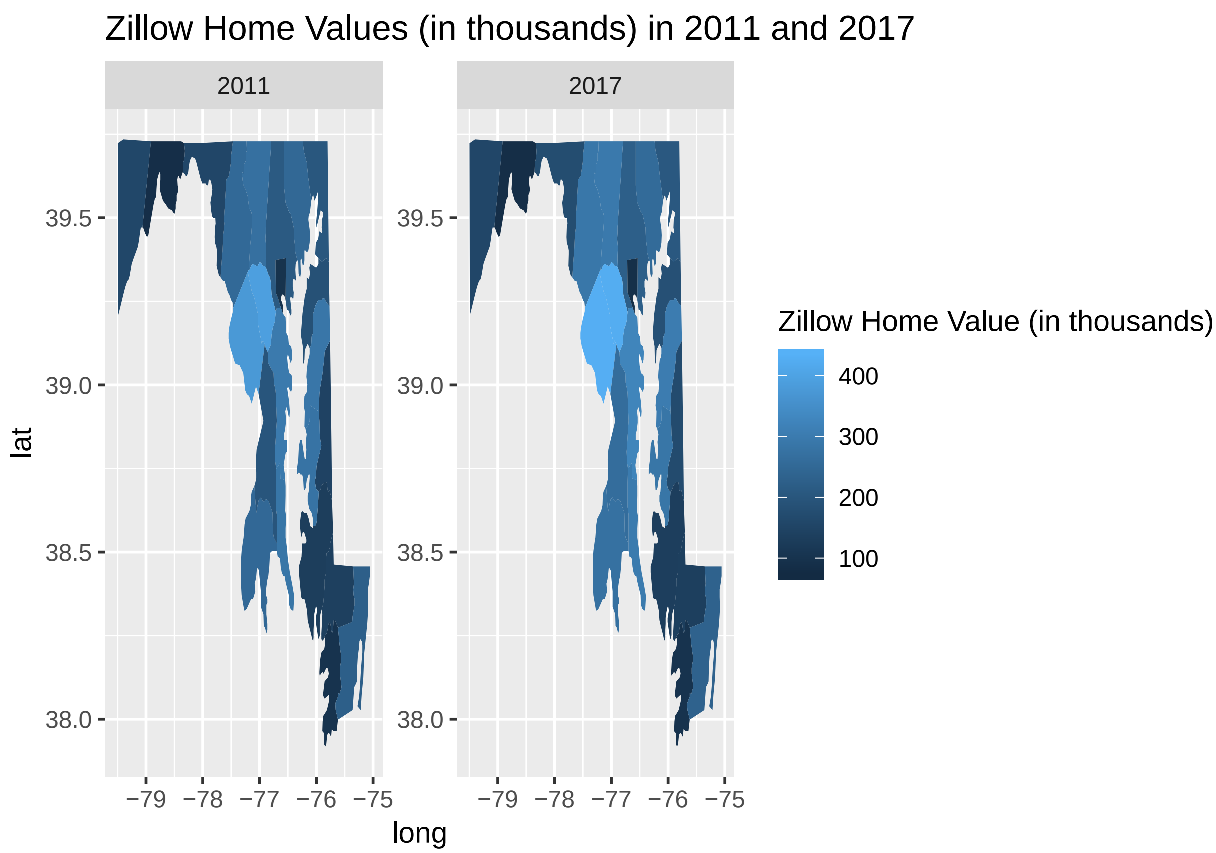


Results

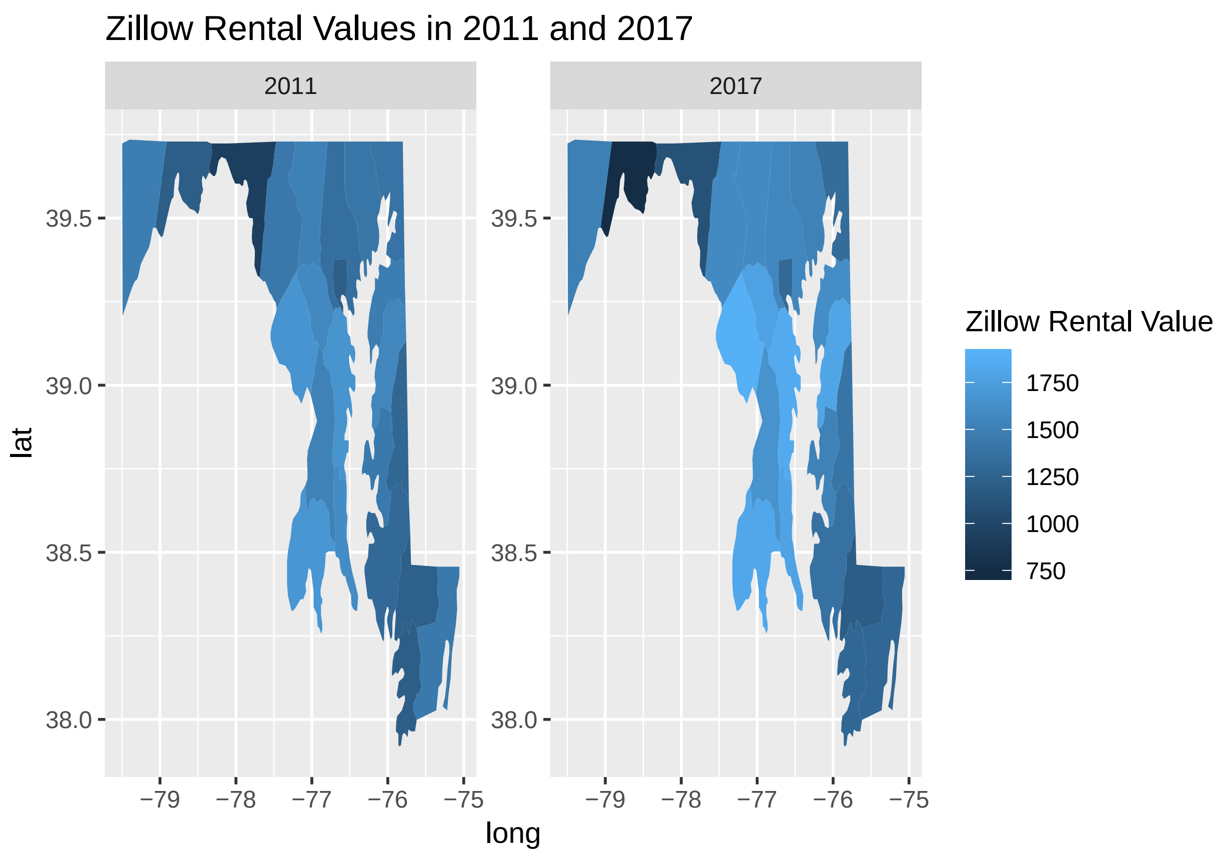
H2: This effect will be more pronounced in urban than in rural counties

Baltimore City can be seen on these maps as a small, square county at roughly (lat -76.5, long 39.3). These maps suggest that Baltimore, the urban center of this metro region, has lower home values and higher rents relative to home values than the surrounding rural and suburban counties do. The results of this visualization lend support to Hypothesis 2.

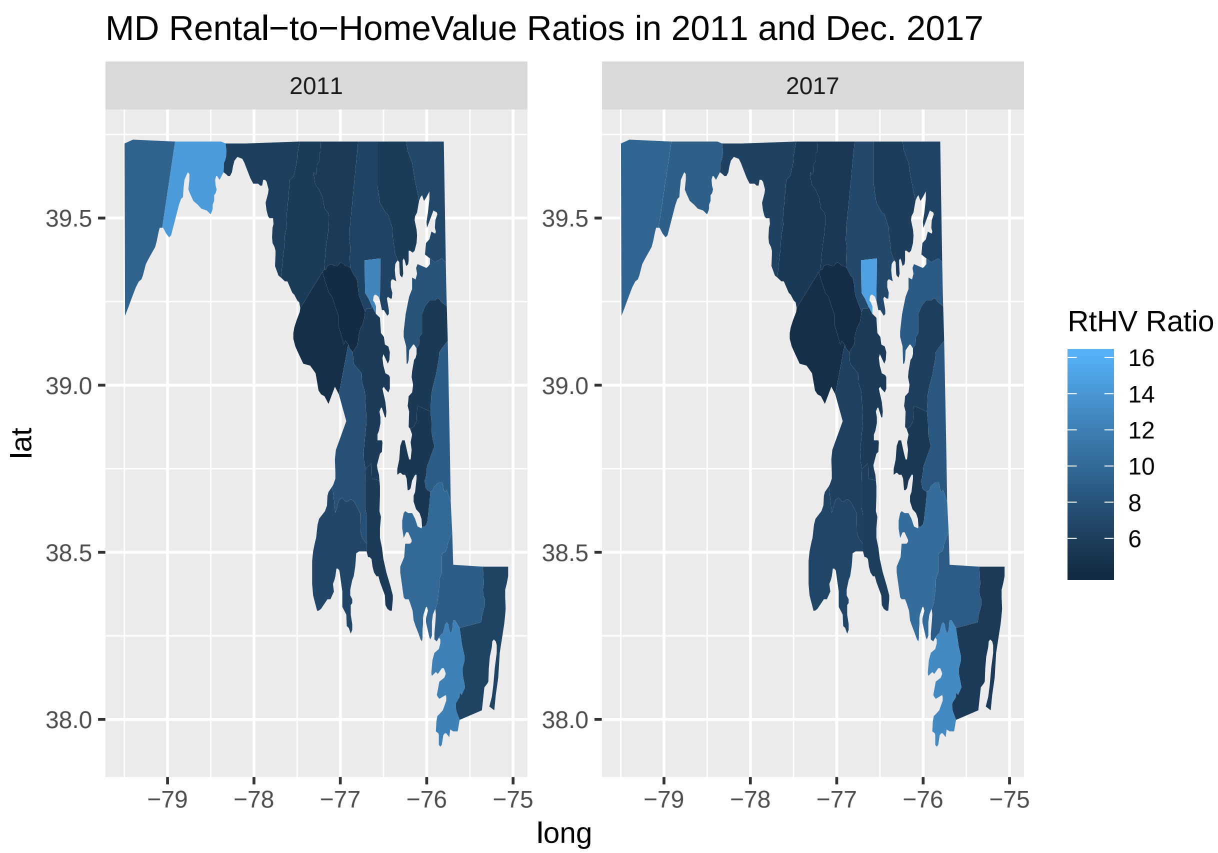
```
#map filling MD counties by average home values in 2017
combinedMD %>%
    mutate(subregion = gsub(" County", "", RegionName)) %>%
    mutate(subregion = tolower(subregion)) %>%
    right_join(md_map, by = "subregion") %>%
    filter(year.x %in% c("2011", "2017")) %>%
    ggplot() +
      geom_polygon(aes(x = long, y = lat, group = subregion, fill = (HomeValue/1000))) +
      facet_wrap(~year.x, scale = "free") +
      ggtitle("Zillow Home Values (in thousands) in 2011 and 2017") +
      scale_fill_continuous(name = "Zillow Home Value (in thousands)")
```




```r
#map filling MD counties by average rental values in 2011
combinedMD %>%
    mutate(subregion = gsub(" County", "", RegionName)) %>%
    mutate(subregion = tolower(subregion)) %>%
    right_join(md_map, by = "subregion") %>%
    filter(year.x %in% c("2011", "2017")) %>%
    ggplot() +
      geom_polygon(aes(x = long, y = lat, group = subregion, fill = RentalValue)) +
      facet_wrap(~year.x, scale = "free") +
      ggtitle("Zillow Rental Values in 2011 and 2017") +
      scale_fill_continuous(name = "Zillow Rental Value")
```




```
#map filling MD counties by average rental values in 2011
combinedMD %>%
    mutate(hrv_ratio = (RentalValue/(HomeValue/1000))) %>%
    mutate(subregion = gsub(" County", "", RegionName)) %>%
    mutate(subregion = tolower(subregion)) %>%
    right_join(md_map, by = "subregion") %>%
    filter(year.x %in% c("2011", "2017")) %>%
    ggplot() +
      geom_polygon(aes(x = long, y = lat, group = subregion, fill = hrv_ratio)) +
      facet_wrap(~year.x, scale = "free") +
      ggtitle("MD Rental-to-HomeValue Ratios in 2011 and Dec. 2017") +
      scale_fill_continuous(name = "RtHV Ratio")
```

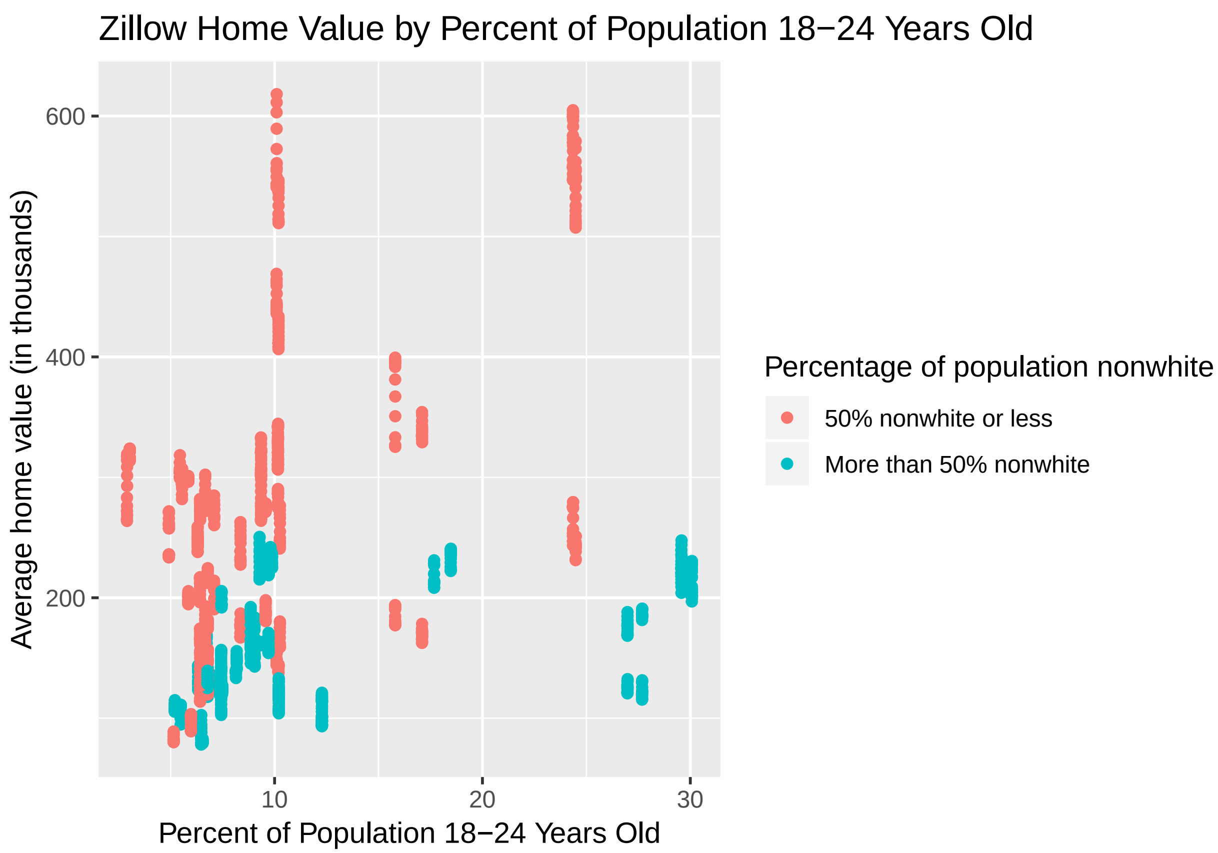


Results

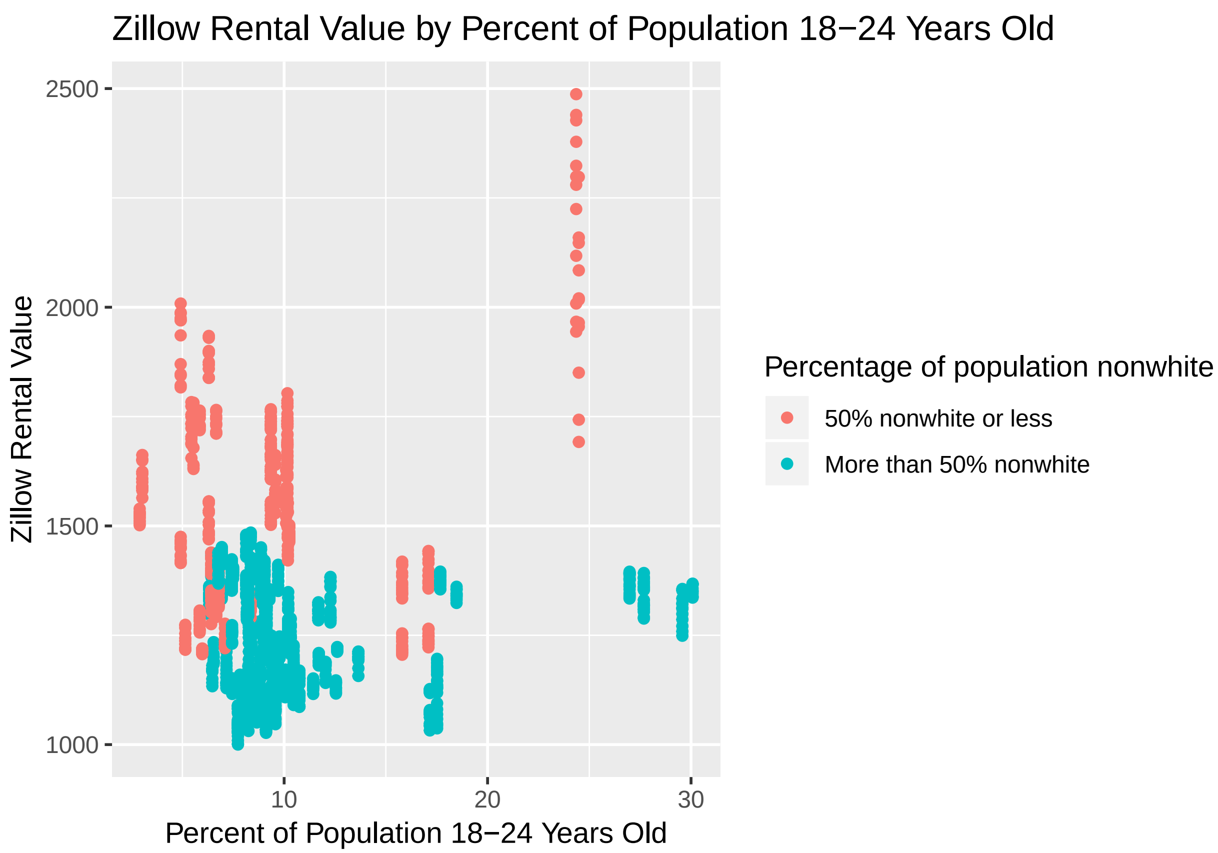
H3: Areas with younger populations will have higher rental values (compared to housing values) than areas with older populations

These scatter plots begin to explore a three-way relationship between nonwhite proportions of the population, young adult (18-24) proportions of the population, and home/rental values or affordability. They nonwhite proportion seems to have a stronger relationship with home/rental values and affordability than young adult population does. Predominantly nonwhite communities have better affordability (which is the proportion of the population spending 30% or more of their income on either rent or a mortgage/housing costs), potentially due to lower rents and home values overall.

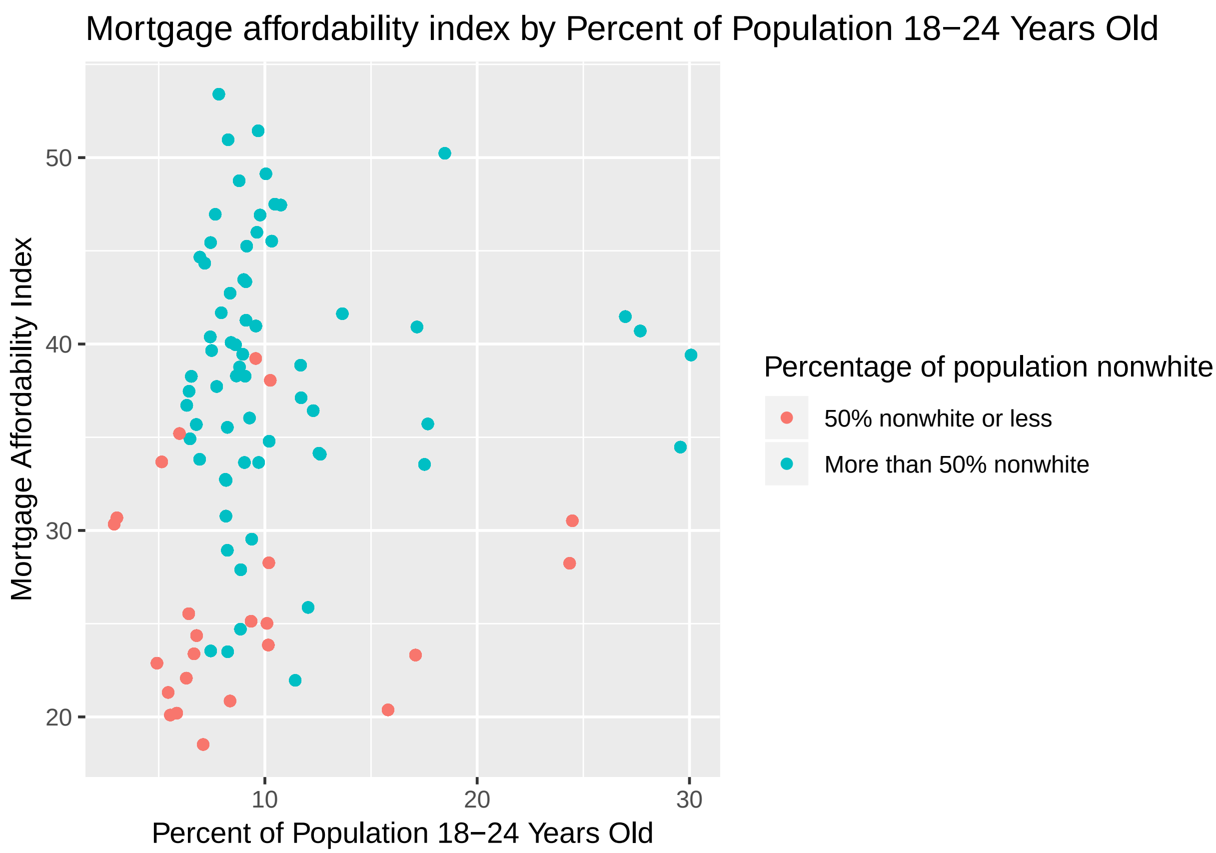
```
#Visualize relationships between age and race make-up of community and average home and rent values in

HVBM %>%
  filter(year %in% c("2016", "2017")) %>%
  mutate(nw_binary = ifelse(nonwhite <= 50, 0, 1)) %>%
  ggplot() +
  geom_jitter(aes(x = percentya, y = HomeValue/1000 , group = RegionName, color = as.factor(nw_binary))
  scale_color_discrete(name = "Percentage of population nonwhite", labels = c("50% nonwhite or less", "
```



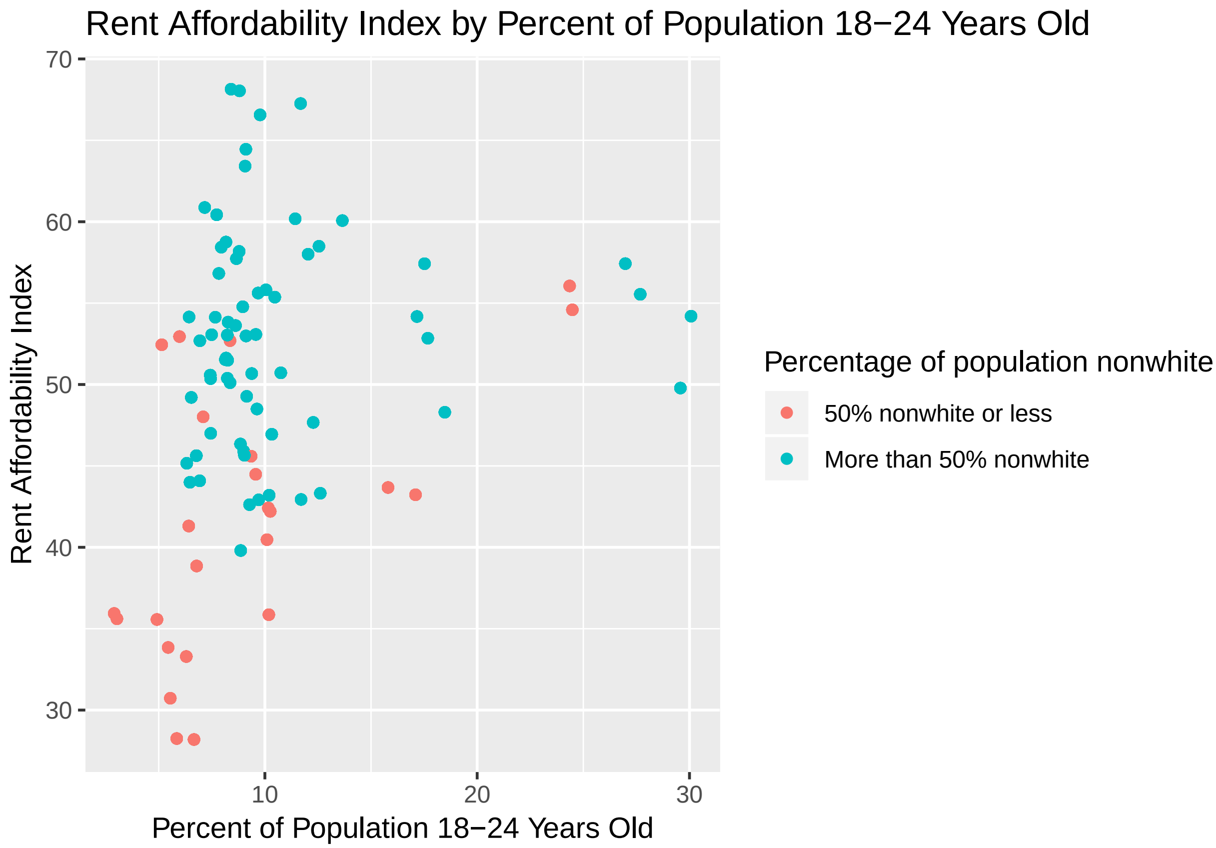

```
RVBM %>%
  filter(year %in% c("2016", "2017")) %>%
  mutate(nw_binary = ifelse(nonwhite <= 50, 0, 1)) %>%
  ggplot() +
  geom_jitter(aes(x = percentya, y = RentalValue , group = RegionName, color = as.factor(nw_binary))) +
  ggtitle("Zillow Rental Value by Percent of Population 18-24 Years Old") +
  xlab("Percent of Population 18-24 Years Old") +
  ylab("Zillow Rental Value") +
  scale_color_discrete(name = "Percentage of population nonwhite", labels = c("50% nonwhite or less", "M
```




```
RVBM %>%
  filter(year %in% c("2016", "2017")) %>%
  mutate(nw_binary = ifelse(nonwhite <= 50, 0, 1)) %>%
  ggplot() +
  geom_jitter(aes(x = percentya, y = ai_mortgage , group = RegionName, color = as.factor(nw_binary))) +
  ggtitle("Mortgage affordability index by Percent of Population 18-24 Years Old") +
  xlab("Percent of Population 18-24 Years Old") +
  ylab("Mortgage Affordability Index") +
  scale_color_discrete(name = "Percentage of population nonwhite", labels = c("50% nonwhite or less", "M
```




```
RVBM %>%
  filter(year %in% c("2016", "2017")) %>%
  mutate(nw_binary = ifelse(nonwhite <= 50, 0, 1)) %>%
  ggplot() +
  geom_jitter(aes(x = percentya, y = ai_rent, group = RegionName, color = as.factor(nw_binary))) +
  ggtitle("Rent Affordability Index by Percent of Population 18-24 Years Old") + xlab("Percent of Popula
  ylab("Rent Affordability Index") +
  scale_color_discrete(name = "Percentage of population nonwhite", labels = c("50% nonwhite or less", "
```

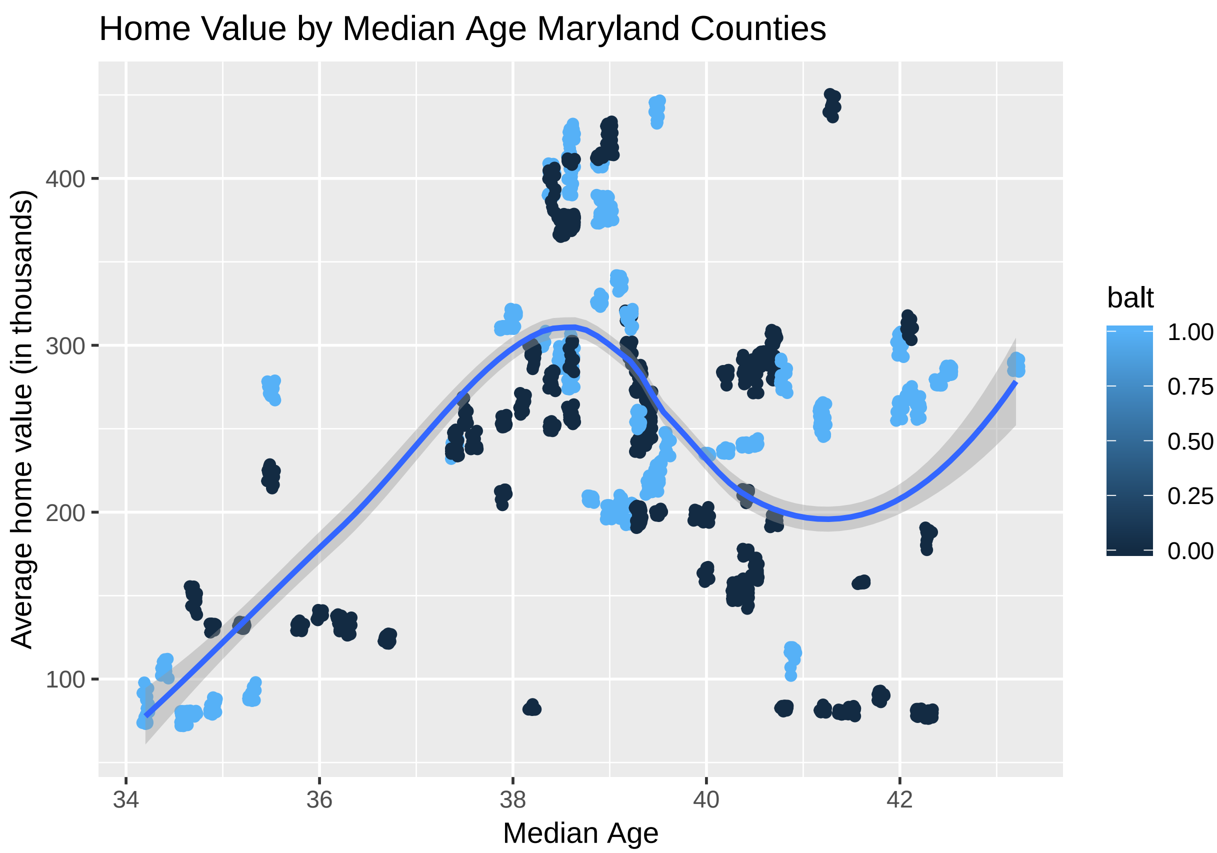


Results

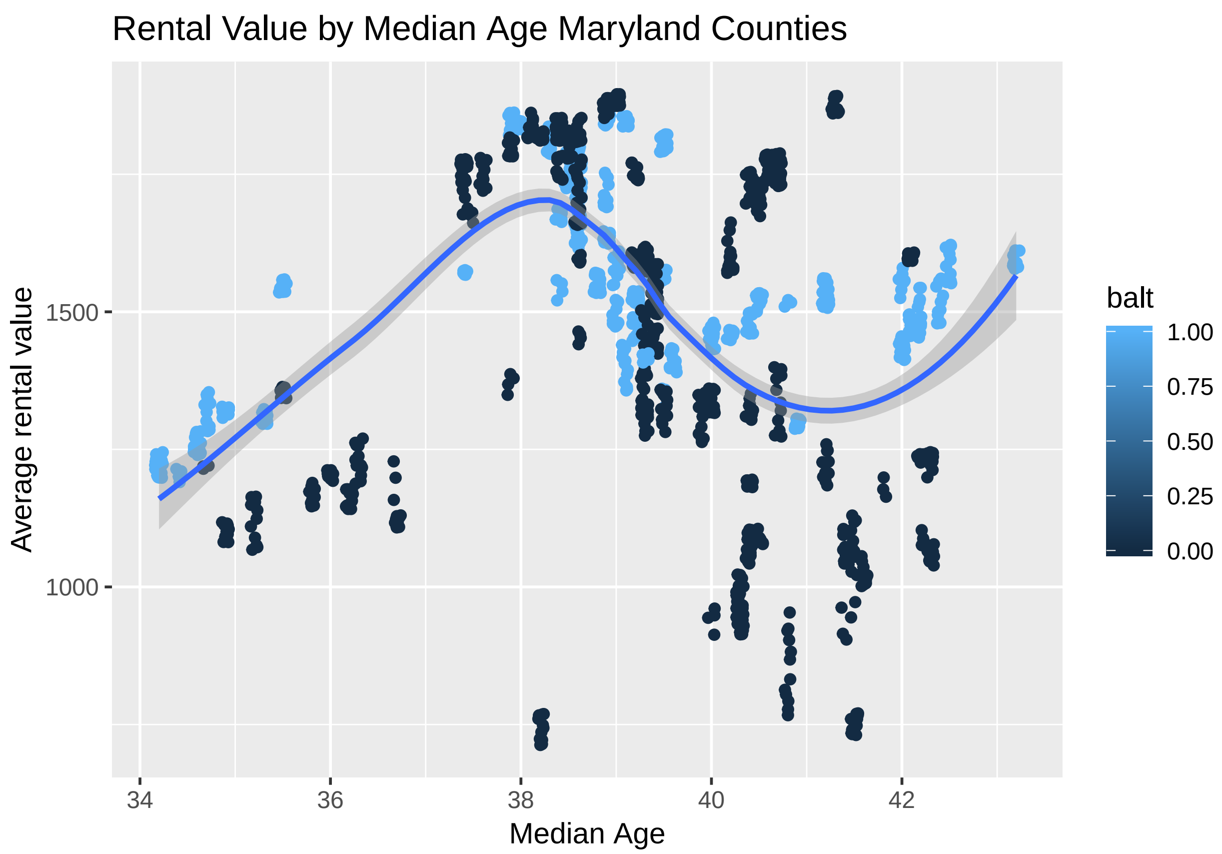
H3 continued

Plotting the data from the ACS reveals that Maryland counties with younger populations generally have lower value homes than those with older populations. This effect is present in the rental values data as well. Looking just at the Baltimore metropolitan area, this trend is even more exaggerated. Moreover, Baltimore City is shown to have a much younger population and much lower average housing values and average rental values than its suburbs. While BNIA did not provide median age for each neighborhood, we can look at the percentage of young people in each neighborhood to understand the relationship between age and price within Baltimore City's neighborhoods. Within the City itself, prices are relatively uniform, though neighborhoods with fewer young people seem to be more expensive both in terms of housing and in terms of renting (except for a few very expensive neighborhoods looked near private universities).

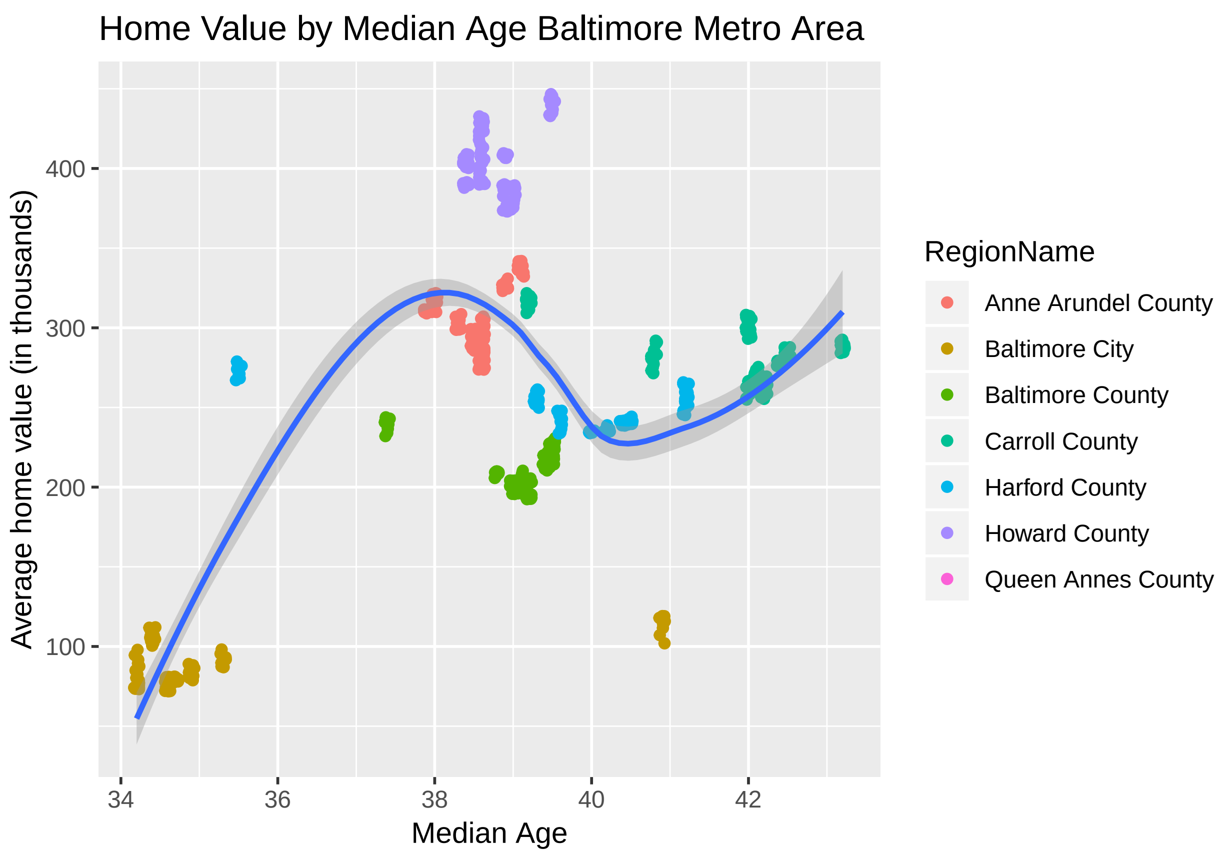
```
HVMD$balt <- as.numeric(HVMD$Metro == "Baltimore-Columbia-Towson")
RVMD$balt <- as.numeric(RVMD$Metro == "Baltimore-Columbia-Towson")

HVMD %>%
  ggplot() +
  geom_jitter(aes(x = medianage, y = HomeValue/1000 , group = RegionName, color = balt))  + geom_smooth
```



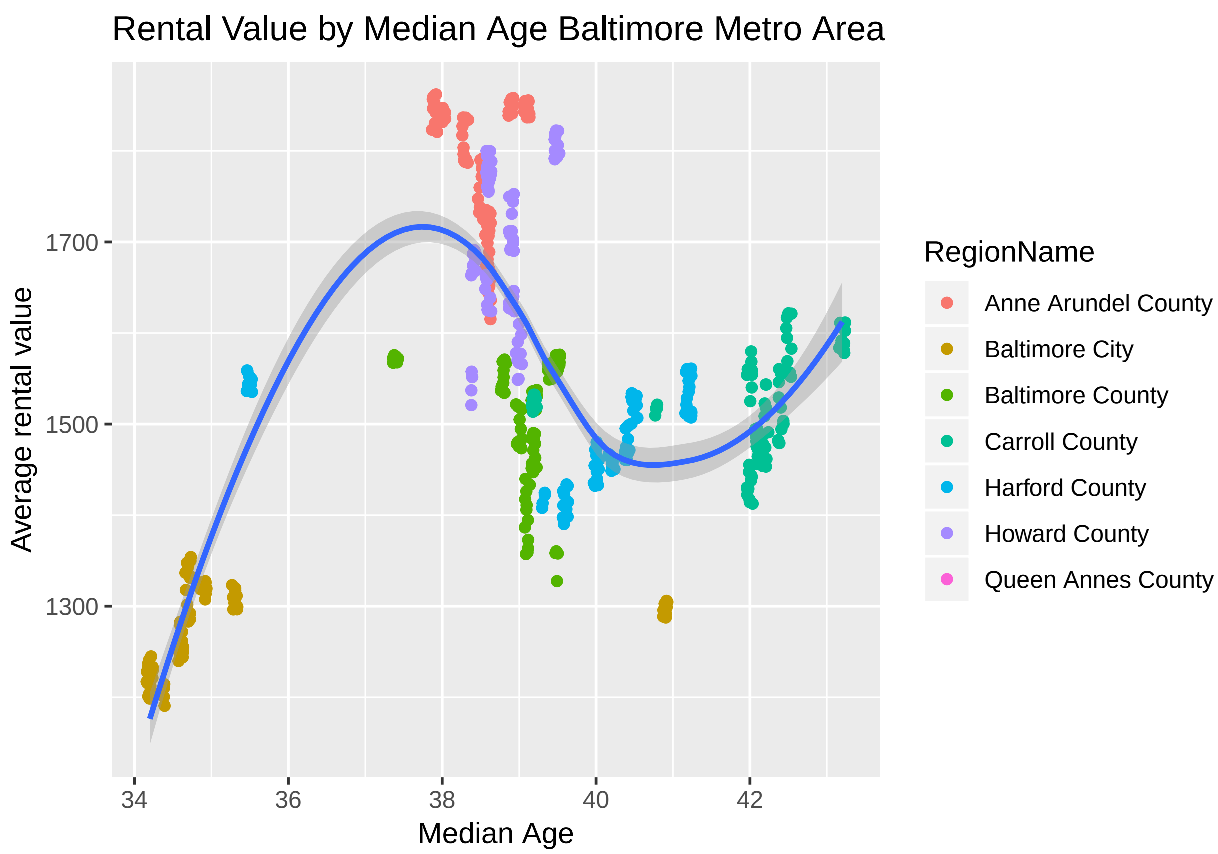

```
RVMD %>%
  ggplot() +
  geom_jitter(aes(x = medianage, y = RentalValue , group = RegionName, color = balt)) + geom_smooth(aes
```



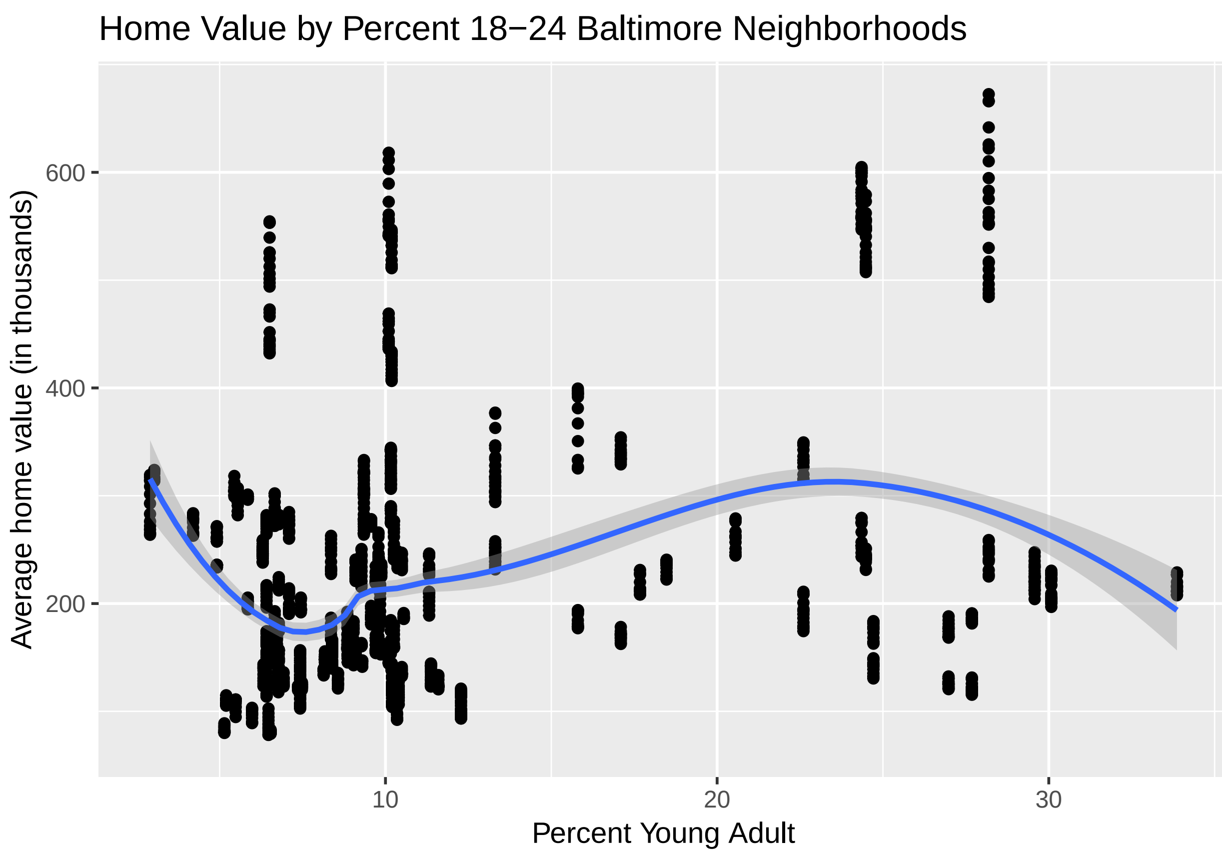

```
HVMD %>%
  filter(Metro == "Baltimore-Columbia-Towson") %>%
  ggplot() +
  geom_jitter(aes(x = medianage, y = HomeValue/1000 , group = RegionName, color = RegionName)) + geom_sm
```



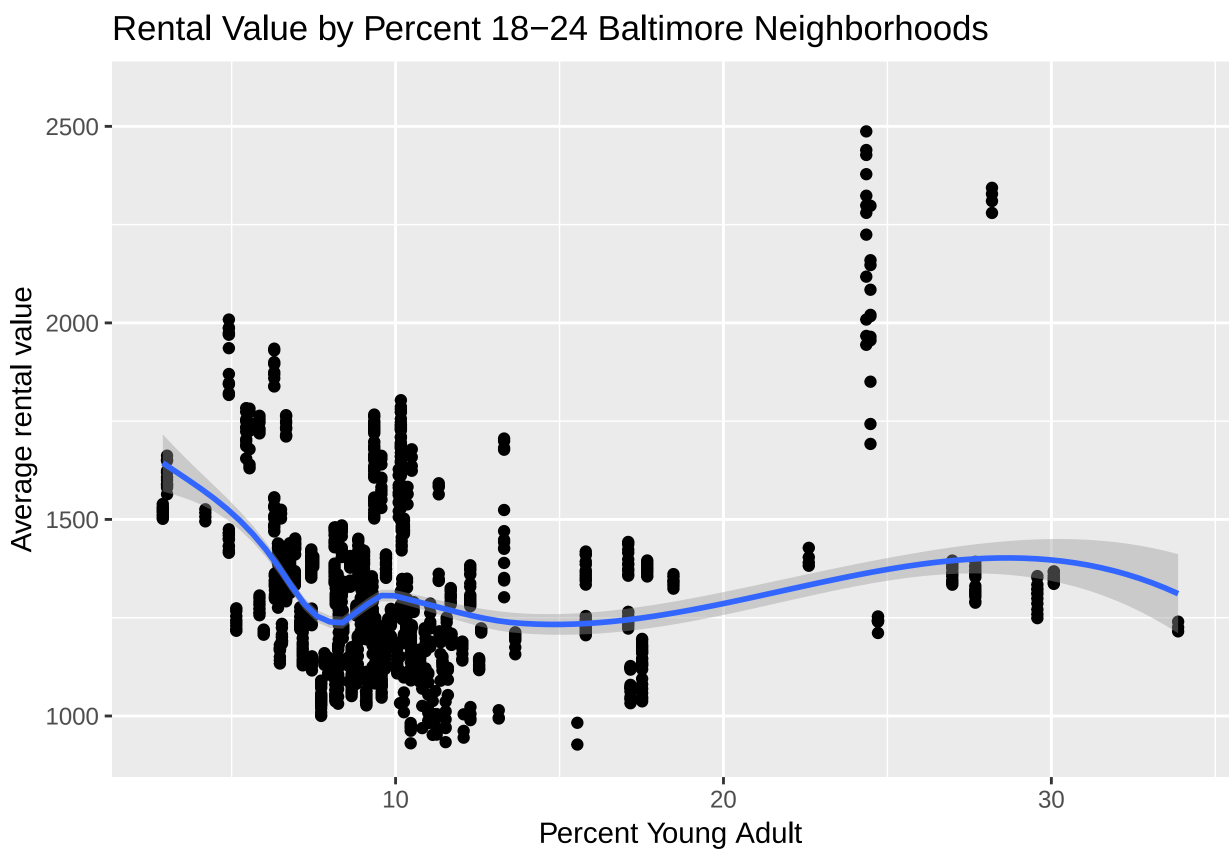

```
RVMD %>%
  filter(Metro == "Baltimore-Columbia-Towson") %>%
  ggplot() +
  geom_jitter(aes(x = medianage, y = RentalValue , group = RegionName, color = RegionName)) + geom_smoot
```




```
HVBM%>%
  ggplot() +
  geom_jitter(aes(x = percentya , y = HomeValue/1000 , group = RegionName)) + geom_smooth(aes(x = percen
```




```r
RVBM%>%
  ggplot() +
  geom_jitter(aes(x = percentya , y = RentalValue, group = RegionName)) + geom_smooth(aes(x = percentya
```

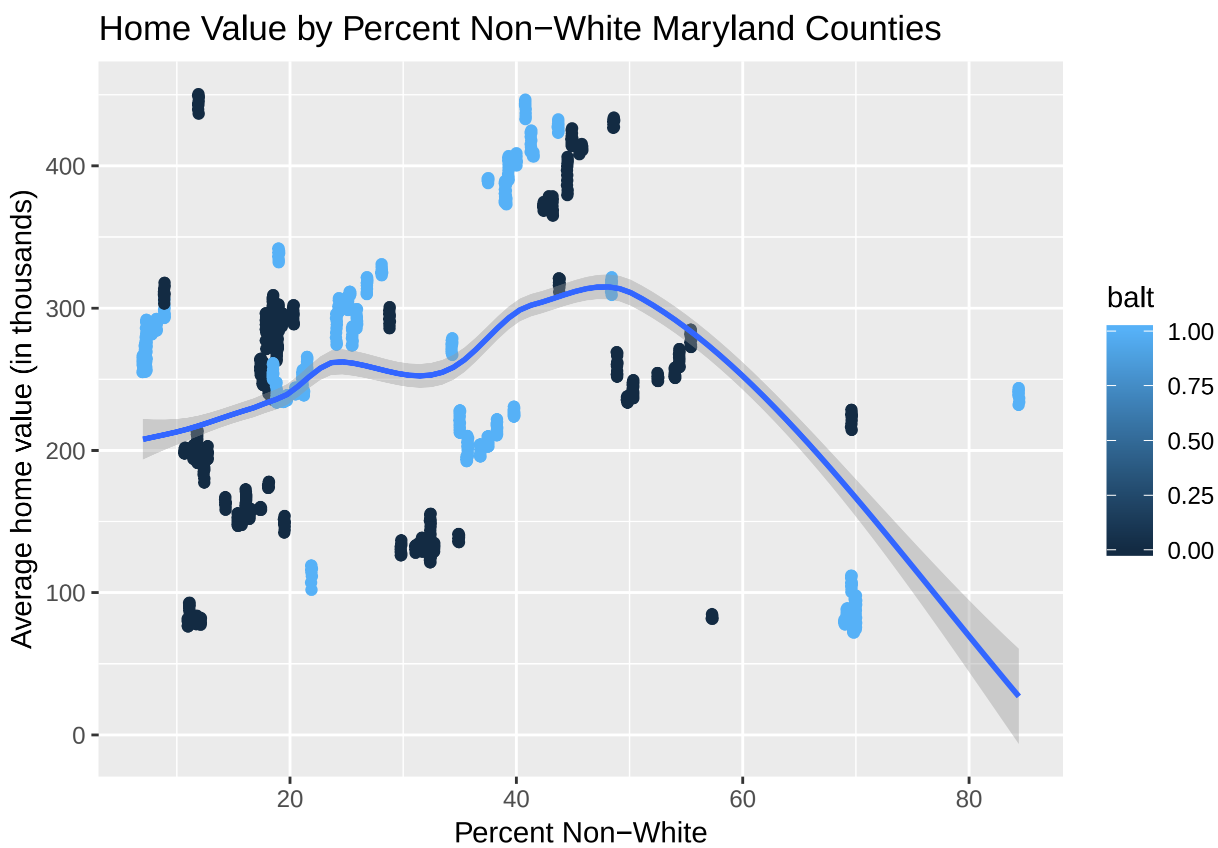


#Results #H4: Increases over time in home and rental values will be associated with decreases in the proportion of non-white residents

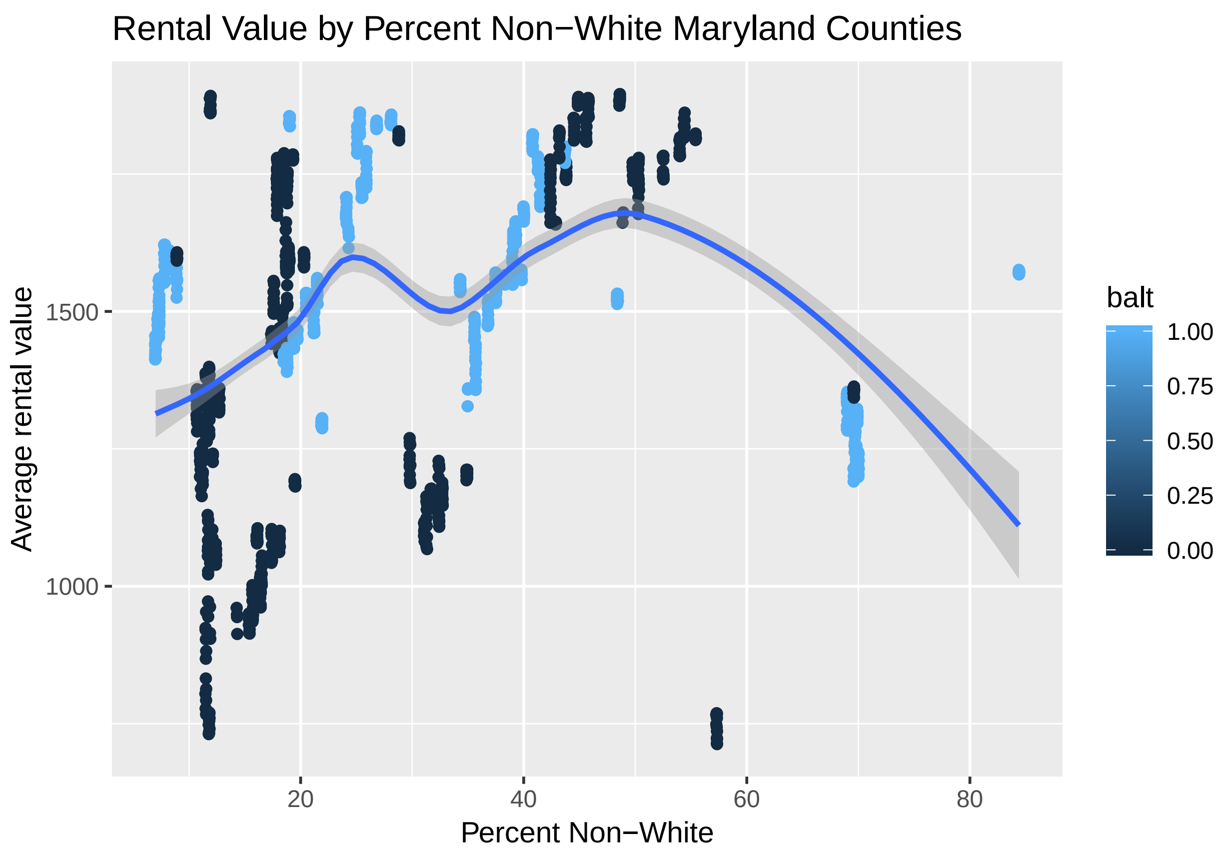
When first looking at a plot of all Maryland counties with the available housing and demographic data, it does not appear that there is a clear relationship between percent non-white and home and rental value. However, when looking at the Baltimore Metropolitan area specifically, it becomes very clear that Baltimore City varies dramatically between its surrounding counties, both in terms of demographics and average home and rental values. Baltimore City is significantly less white than the surrounding counties and has much lower housing values.

To better illustrate this disparity, we can compare housing values in Baltimore City to Baltimore County. Baltimore City is around 70% non-white, while Baltimore County, the location of Baltimore City's more populous suburbs, is less than 10% non-white—a huge racial gap! While there is a significant gap in home values between Baltimore City and County, the gap appears to have remained constant since 2010. However, not only does is there a similar gap in rental values between Baltimore County and Baltimore City, but the gap seems to have continuously widened. This widening gap in rental prices may serve to make it more difficult for the younger and less white population of Baltimore City to move to areas like Baltimore County. Within Baltimore City itself, this trend is also present and the neighborhoods with the highest home and rental values are much whiter than most of the poorer neighborhoods in the city

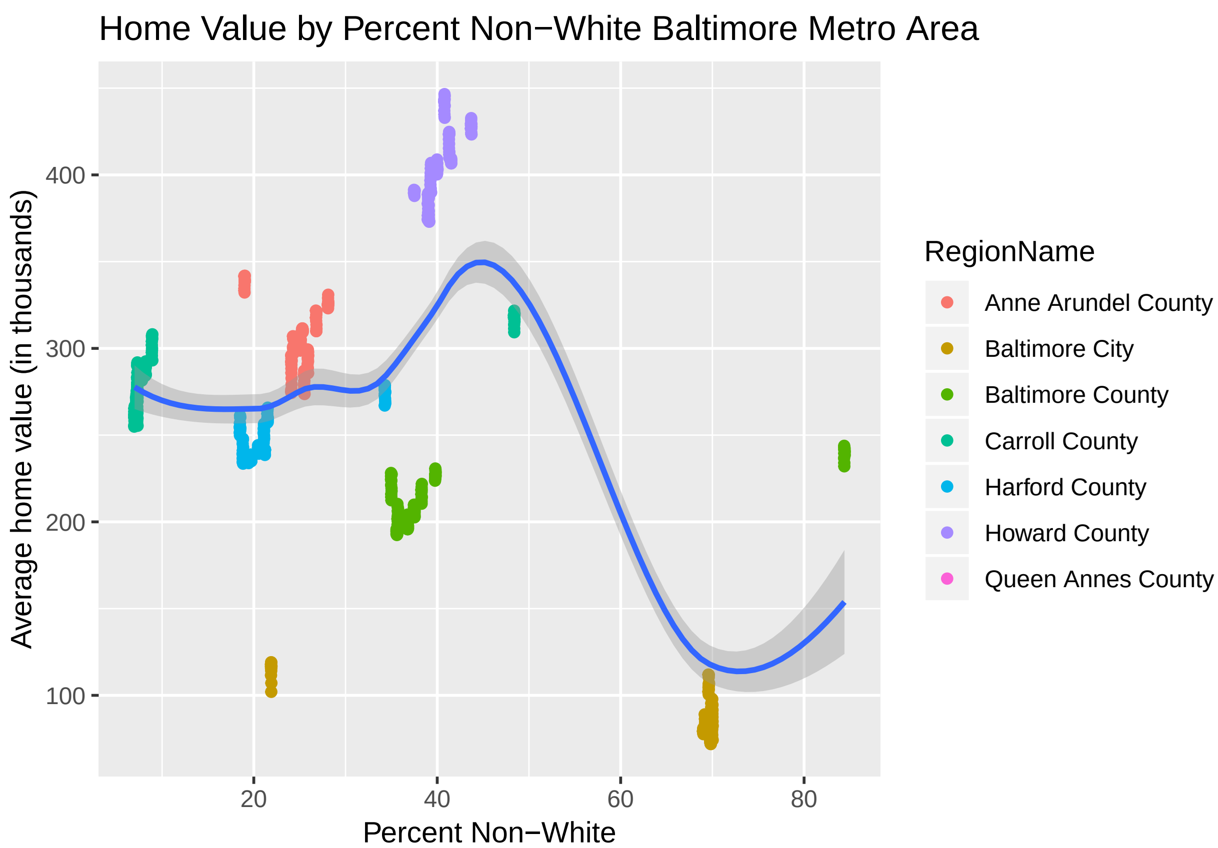
```
HVMD %>%
  ggplot() +
  geom_jitter(aes(x = nonwhite, y = HomeValue/1000 , group = RegionName, color = balt))  + geom_smooth(a
```



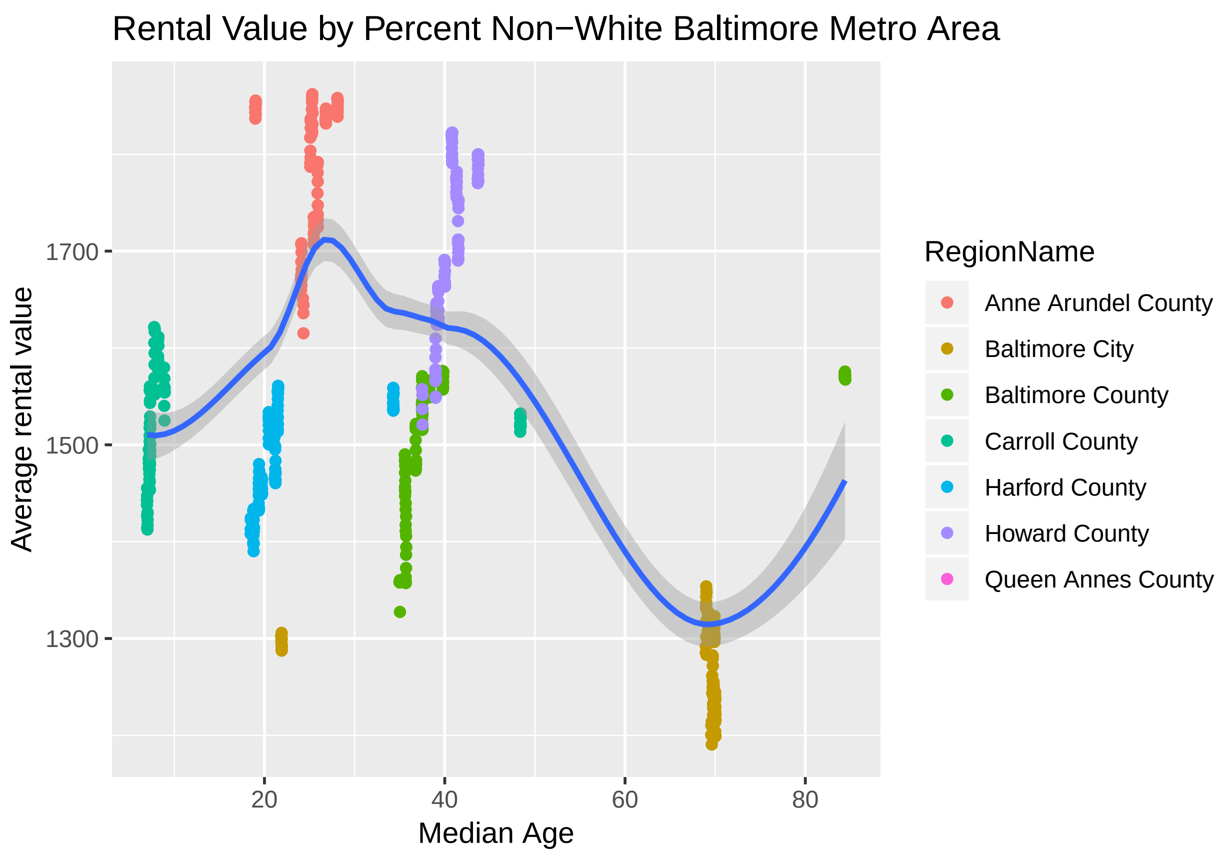

```r
RVMD %>%
  ggplot() +
  geom_jitter(aes(x = nonwhite, y = RentalValue , group = RegionName, color = balt)) + geom_smooth(aes(
```



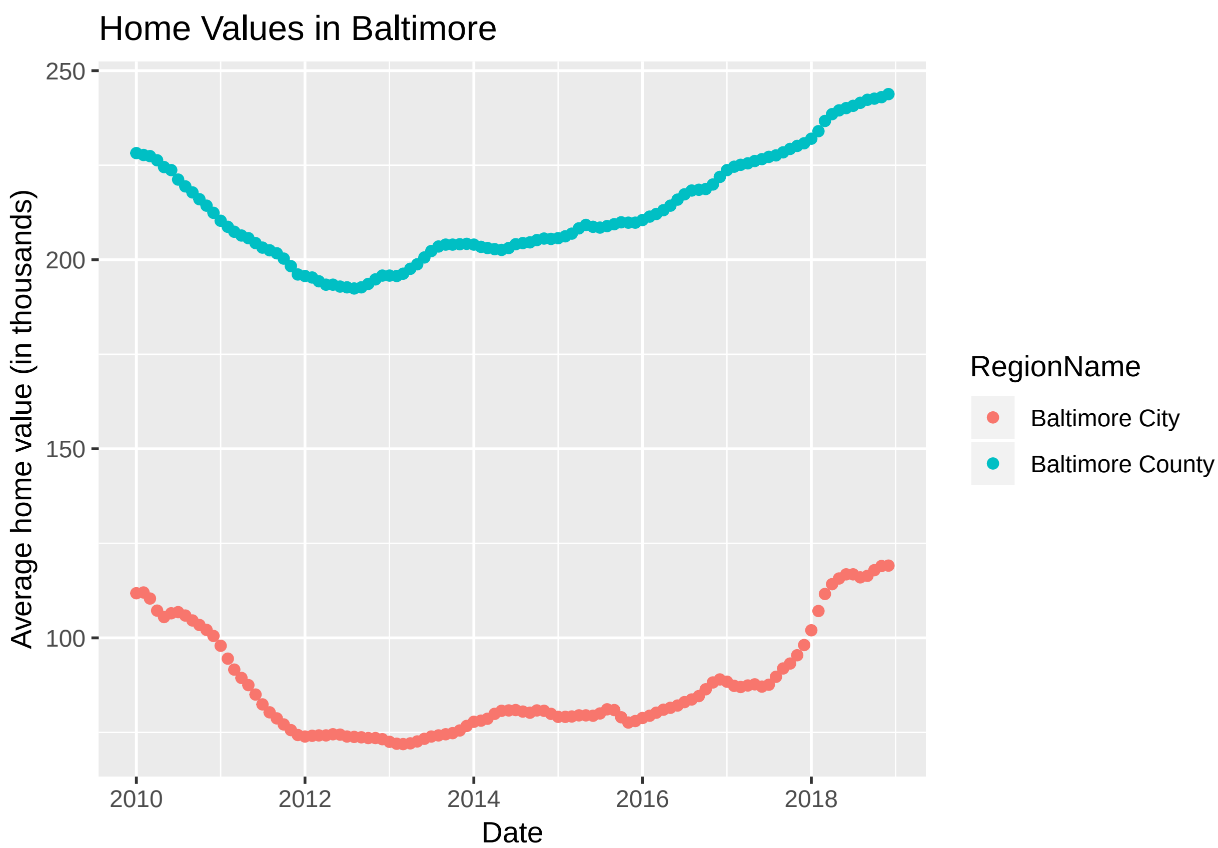

```
HVMD %>%
  filter(Metro == "Baltimore-Columbia-Towson") %>%
  ggplot() +
  geom_jitter(aes(x = nonwhite, y = HomeValue/1000 , group = RegionName, color = RegionName)) + geom_smo
```



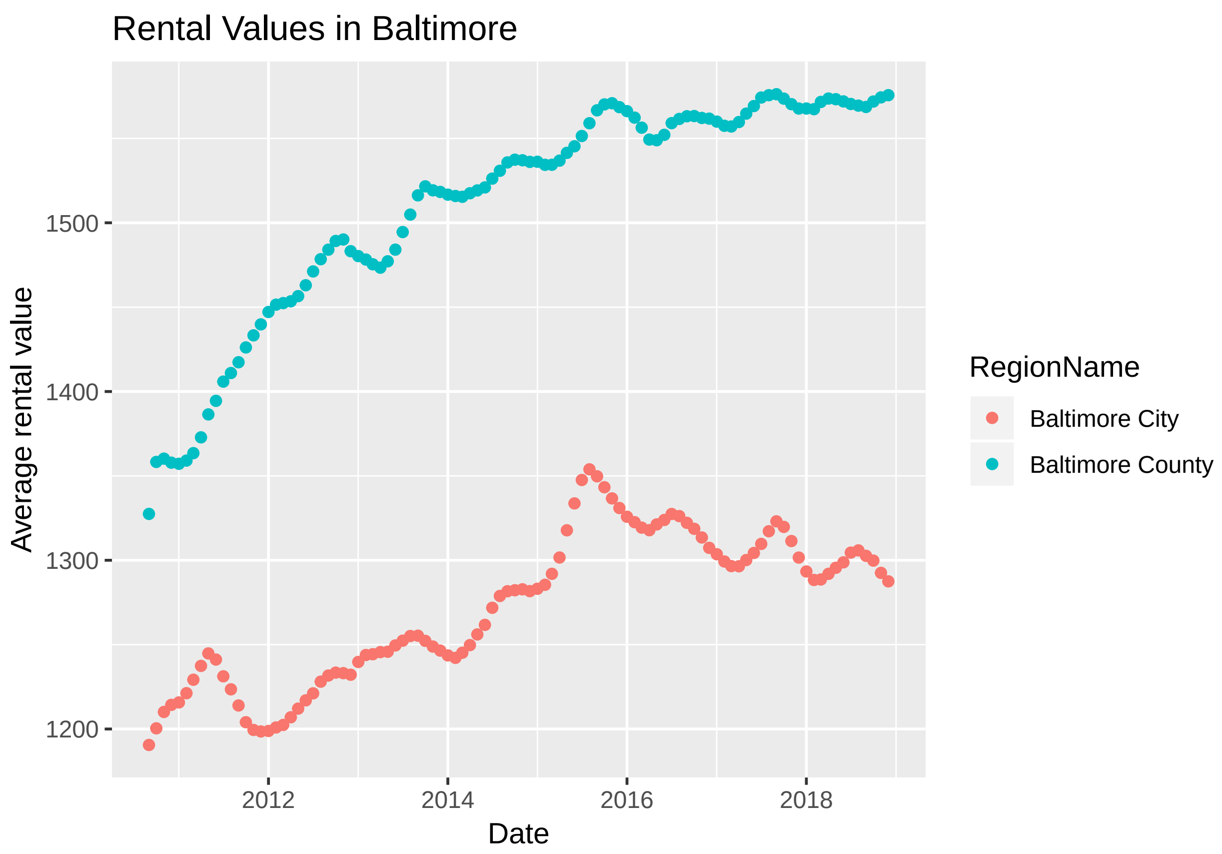

```
RVMD %>%
  filter(Metro == "Baltimore-Columbia-Towson") %>%
  ggplot() +
  geom_jitter(aes(x = nonwhite, y = RentalValue , group = RegionName, color = RegionName)) + geom_smoot
```



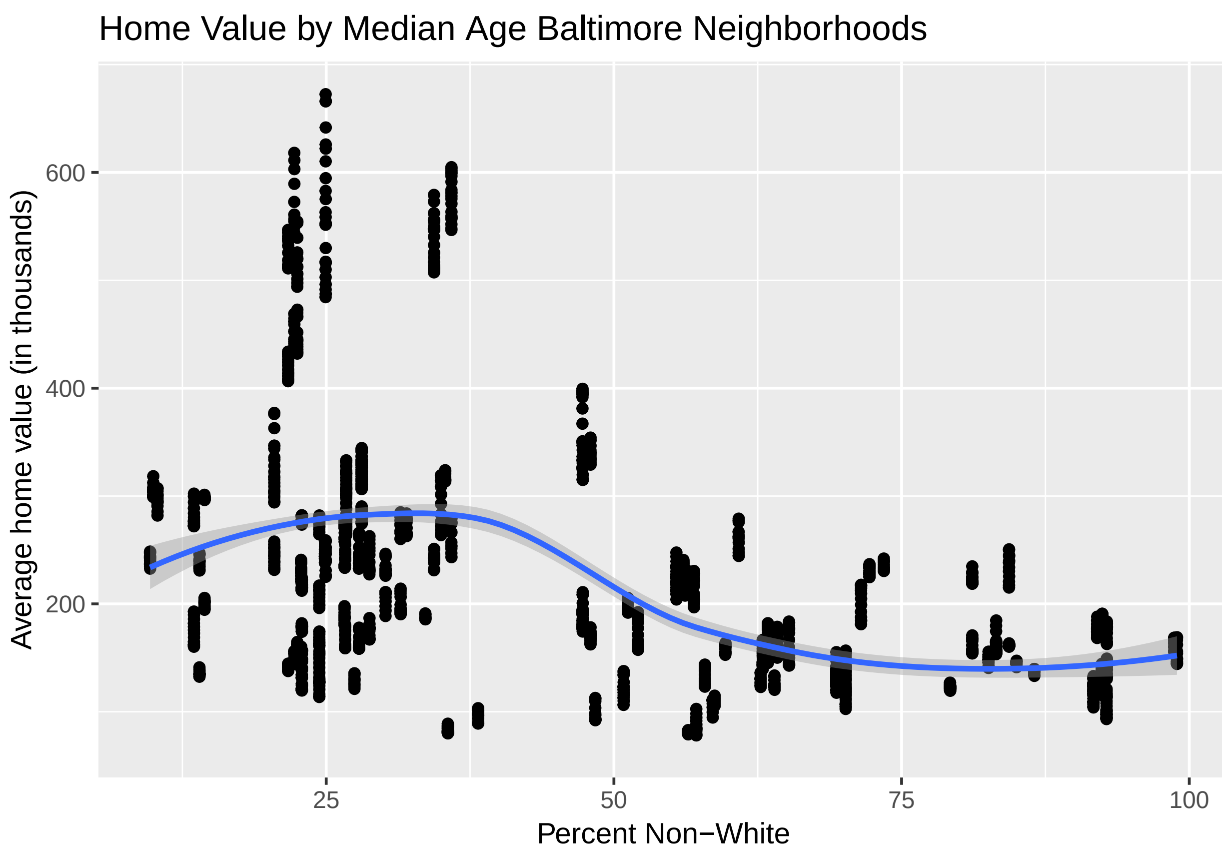

```
HVMD %>%
  filter(RegionName == "Baltimore County" |RegionName == "Baltimore City" ) %>%
  ggplot() +
  geom_point(aes(x = as.Date(paste(date,"-01",sep="")), y = HomeValue/1000, color = RegionName)) + ggti
```



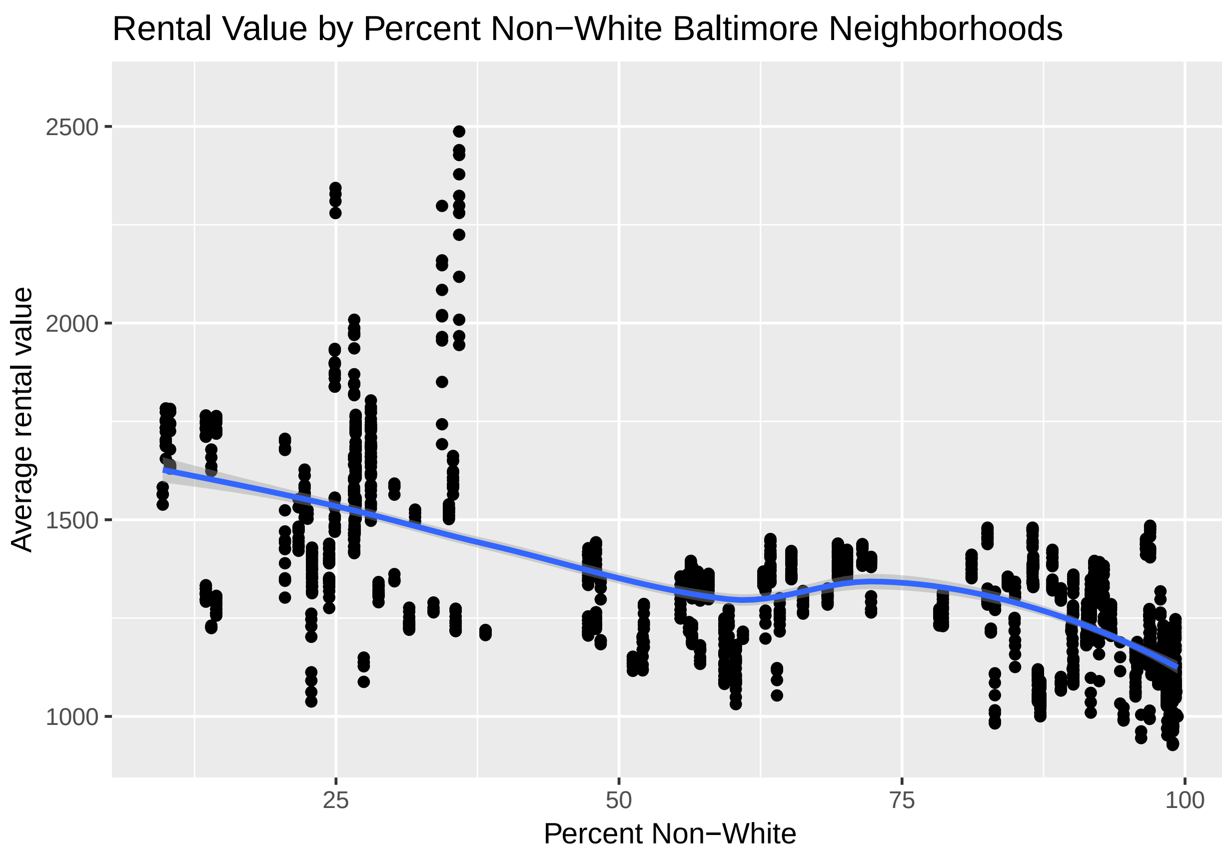

```
RVMD %>%
  filter(RegionName == "Baltimore County" |RegionName == "Baltimore City" ) %>%
  ggplot() +
  geom_point(aes(x = as.Date(paste(date,"-01",sep="")), y = RentalValue, color = RegionName)) + ggtitle
```




```
HVBM%>%
  ggplot() +
  geom_jitter(aes(x = nonwhite, y = HomeValue/1000 , group = RegionName)) + geom_smooth(aes(x = nonwhit
```




```
RVBM%>%
  ggplot() +
  geom_jitter(aes(x = nonwhite, y = RentalValue, group = RegionName)) + geom_smooth(aes(x = nonwhite, y
```



Discussion

While Zillow's housing data proved useful for mapping at the state/county level, and showed similar distributions across communities to the BNIA data, there were a number of disadvantages to working with this data. There were many missing values that could interrupt trend lines and produced blank areas when mapping – in particular this is visible in the map of HousingValue during 2010, where the lower-home value, less-white portions of the city are largely missing. This suggests that the potential for bias in the Zillow data may increase the further back the data goes – this makes sense, as one would expect their database to become more robust over time, as use of the site has increased and techniques for data mining have improved.

Another major issue we encountered with the Zillow datasets is the lack of standardized geographic identifiers for merging with other data and mapping at meaningful community levels – Zillow data had to be merged to both ACS and BNIA data by county or neighborhood name, and these text-based identifiers proved to be significantly more labor-intensive to preprocess. Zillow's neighborhood names also had significant alignment issues with the only publicly-available Baltimore neighborhood shapefile, leading to large blank areas when mapping was attempted.

Neighborhood names in general proved to be problematic to work with – in fact, we could not obtain demogrpahic data for areas smaller than the Community Statistical Areas, which we were able to map to neighborhoods using BNIA's provided geographic crosswalk, but encountered many problems in merging due to misspellings, neighborhoods with multiple names, neighborhoods that are sometimes counted as part of a larger surrounding neighborhood, and abbreviations.

Finally, we found it difficult to locate comparable measures to Zillow's home value and median rental value indexes, making it difficult to assess the validity of these measures. We do have some evidence of their validity at least in terms of comparing different geographic areas, as the Zillow data displayed similar trends to the

BNIA data when mapped onto Baltimore city by neighborhood/CSA. However, it would not be possible to make a general statement about the validity of these measures based

This exploration of exploring housing and demographic relationships effectively illustrates the difficulties, limitations, and extra processing burden of working with open access data, as well as the advantages. Zillow data allowed us to compare market estimates to survey-based estimates and visualize seasonal and other fine-grain temporal changes in housing & rental markets, and provided a more direct estimate of rental markets than was available from any of our other data sources.

References

Brummet, Q. and Reed D., (2019), The Effects of Gentrification on the Well-Being and Opportunity of Original Resident Adults and Children. FRB of Philadelphia Working Paper No. 19-30.

Capps, K., (2019, July 16), The Hidden Winners in Neighborhood Gentrification. Retrieved from https://www.citylab.com/equity/2019/07/gentrification-effects-neighborhood-data-economic-statistics/594064/.

Hoffower, H., (2019, July 29), Millennials aren't buying homes, and it might not be because they can't afford them: Some actually prefer to rent instead. Retrieved from https://www.businessinsider.com/millennials-renting-homes-instead-of-buying-2019-7.

Hoynes, H.W. and McFadden, D. (1994), The Impact of Demographics on Housing and Non-Housing Wealth in the United States. NBER Working Paper No. 4666.

Kasakove, S., (2019, September 25), AOC's Plan to Decommodify Housing. Retrieved from https://www.thenation.com/article/aoc-sanders-housing-rent/.

Meehan, S., (2019, March 20), Baltimore among nation's most gentrified cities, study shows. Retrieved from https://www.baltimoresun.com/maryland/baltimore-city/bs-md-ci-gentrification-study-20190319-story.html.

Squires, G. and Kubrin, C., (2005), Privileged Places: Race, Uneven Development, and the Geography of Opportunity in Urban America. Urban Studies, 42.