

Tutorial 7: Maps, 2 of 2

Leah Brooks

March 24, 2019

Welcome back to maps! In today's second of two mapping tutorials, we learn how to make

- choropleth maps
- histogram legends for those maps
- dot density maps

We also learn some useful R programming, including

- `gather()`: the opposite of `spread()`, which we already learned
- how to annotate a chart in R

Unlike all the tutorials so far in this class, I did not give this to Jill in advance to proof. All errors are my own. I also encountered a number of odd error messages while working on this tutorial, frequently things like "reached elapsed time limit." These messages didn't seem to inhibit the program, however.

A. Load Packages and Data

We are using a set of packages that you should have already used in the past. Here I am loading them with `require()` which installs the package if you do not already have it.

```
require(sf)
```

```
## Loading required package: sf
## Linking to GEOS 3.6.1, GDAL 2.2.3, proj.4 4.9.3
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(dplyr)
```

```
## Loading required package: dplyr
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
require(tidyr)
```

```
## Loading required package: tidyr
```

Like the previous map tutorial class, we'll use crime data. See the [previous tutorial](#), section D.1. to re-load these data if you need.

```

# we'll use the crime data again
c2018 <- st_read("H:/pppa_data_viz/2019/tutorial_data/lecture05/Crime_Incidents_in_2018/Crime_Incidents

## Reading layer `Crime_Incidents_in_2018' from data source `H:\pppa_data_viz\2019\tutorial_data\lectur
## Simple feature collection with 33645 features and 23 fields
## geometry type: POINT
## dimension: XY
## bbox: xmin: -77.11232 ymin: 38.81467 xmax: -76.91002 ymax: 38.9937
## epsg (SRID): 4326
## proj4string: +proj=longlat +datum=WGS84 +no_defs

```

To make these data more manageable, we limit ourselves to the very serious crimes of burglary and homicide.

```

# make smaller dataframe
vc2018 <- c2018[which(c2018$OFFENSE == "HOMICIDE" | c2018$OFFENSE == "BURGLARY"),]
names(vc2018)

```

```

## [1] "CCN" "REPORT_DAT" "SHIFT" "METHOD" "OFFENSE"
## [6] "BLOCK" "XBLOCK" "YBLOCK" "WARD" "ANC"
## [11] "DISTRICT" "PSA" "NEIGHBORHO" "BLOCK_GROU" "CENSUS_TRA"
## [16] "VOTING_PRE" "LATITUDE" "LONGITUDE" "BID" "START_DATE"
## [21] "END_DATE" "OBJECTID" "OCTO_RECOR" "geometry"

```

We will do all geographic analysis for today's tutorial at the block group level, so load the block group map you used in Section E of the [previous map tutorial](#).

```

# load block group map for dc only
bg2010 <- st_read("H:/pppa_data_viz/2019/tutorial_data/lecture05/Census_Block_Groups__2010/Census_Block

## Reading layer `Census_Block_Groups__2010' from data source `H:\pppa_data_viz\2019\tutorial_data\lect
## Simple feature collection with 450 features and 54 fields
## geometry type: POLYGON
## dimension: XY
## bbox: xmin: -77.11976 ymin: 38.79165 xmax: -76.9094 ymax: 38.99581
## epsg (SRID): 4326
## proj4string: +proj=longlat +datum=WGS84 +no_defs
names(bg2010)

```

```

## [1] "OBJECTID" "TRACT" "BLKGRP" "GEOID" "P0010001"
## [6] "P0010002" "P0010003" "P0010004" "P0010005" "P0010006"
## [11] "P0010007" "P0010008" "0P000001" "0P000002" "0P000003"
## [16] "0P000004" "P0020002" "P0020005" "P0020006" "P0020007"
## [21] "P0020008" "P0020009" "P0020010" "0P000005" "0P000006"
## [26] "0P000007" "0P000008" "P0030001" "P0030003" "P0030004"
## [31] "P0030005" "P0030006" "P0030007" "P0030008" "0P000009"
## [36] "0P000010" "0P000011" "0P000012" "P0040002" "P0040005"
## [41] "P0040006" "P0040007" "P0040008" "P0040009" "P0040010"
## [46] "0P000013" "0P000014" "0P000015" "0P000016" "H0010001"
## [51] "H0010002" "H0010003" "SHAPE_Leng" "SHAPE_Area" "geometry"

```

B. Create a block-group level shapefile with crime data

To make block-group level maps, we need a block-group level simple feature (the `sf` version of a shapefile) with crime information. What we have now is a block group level simple feature and a crime dataset that is

at the level of the individual crime. So we need to modify the crime dataset so make it at the block group level and then add these aggregated crime data to the simple feature.

We begin by figuring out which crimes are in which block groups. In the interests of full disclosure, this information is actually already in `vc2018`, but I am doing this so you know how to do it when you don't already have this helpful information in hand. We something very similar in the previous map class, so these steps should be somewhat of a review.

To make sure that all we are adding to the crime data is the block group information, we create a new dataframe called `bg2010.small`. This dataframe has just the tract and block group ids (recall, a block group id is the tract id plus the block group id – 7 digits in total). We then intersect the crime data with this block group simple feature. This tells us, for each observation in `vc2018`, in which block group it lies.

```
# find which crimes are in which block groups
# vc2010 actually already has the block group, but i want to be sure you can do this yourself
bg2010.small <- bg2010[,c("TRACT", "BLKGRP")]
cbg <- st_intersection(vc2018, bg2010.small)
```

```
## although coordinates are longitude/latitude, st_intersection assumes that they are planar
## Warning: attribute variables are assumed to be spatially constant
## throughout all geometries
```

```
head(cbg)
```

```
## Simple feature collection with 6 features and 25 fields
## geometry type: POINT
## dimension: XY
## bbox: xmin: -77.0734 ymin: 38.90259 xmax: -77.05759 ymax: 38.91258
## epsg (SRID): 4326
## proj4string: +proj=longlat +datum=WGS84 +no_defs
## CCN REPORT_DAT SHIFT METHOD OFFENSE
## 1826 18064713 2018-04-23T13:04:28.000Z DAY OTHERS BURGLARY
## 11452 18134729 2018-08-14T03:10:42.000Z MIDNIGHT OTHERS BURGLARY
## 26674 18042061 2018-03-15T16:09:21.000Z EVENING OTHERS BURGLARY
## 30745 18026424 2018-02-16T14:31:32.000Z DAY OTHERS BURGLARY
## 30750 18026435 2018-02-16T15:42:03.000Z EVENING OTHERS BURGLARY
## 23472 18136654 2018-08-17T04:38:28.000Z MIDNIGHT OTHERS BURGLARY
## BLOCK XBLOCK YBLOCK WARD ANC
## 1826 3100 - 3199 BLOCK OF K STREET NW 394626 137194 2 2E
## 11452 3036 - 3099 BLOCK OF M STREET NW 394737 137483 2 2E
## 26674 3000 - 3099 BLOCK OF N STREET NW 394778 137665 2 2E
## 30745 2800 - 2899 BLOCK OF PENNSYLVANIA AVENUE NW 395005 137471 2 2E
## 30750 2800 - 2899 BLOCK OF PENNSYLVANIA AVENUE NW 395005 137471 2 2E
## 23472 3700 - 3799 BLOCK OF RESERVOIR ROAD NW 393634 138304 2 2E
## DISTRICT PSA NEIGHBORHO BLOCK_GROU CENSUS_TRA VOTING_PRE LATITUDE
## 1826 2 206 Cluster 4 000100 4 000100 Precinct 5 38.90258
## 11452 2 206 Cluster 4 000100 4 000100 Precinct 5 38.90519
## 26674 2 206 Cluster 4 000100 4 000100 Precinct 5 38.90683
## 30745 2 206 Cluster 4 000100 4 000100 Precinct 5 38.90508
## 30750 2 206 Cluster 4 000100 4 000100 Precinct 5 38.90508
## 23472 2 206 Cluster 4 000201 1 000201 Precinct 6 38.91258
## LONGITUDE BID START_DATE
## 1826 -77.06195 GEORGETOWN 2018-04-23T10:56:35.000Z
## 11452 -77.06068 GEORGETOWN 2018-08-13T22:52:11.000Z
## 26674 -77.06021 <NA> 2018-03-14T03:48:45.000Z
## 30745 -77.05759 GEORGETOWN 2018-02-16T01:25:28.000Z
```

```
## 30750 -77.05759 GEORGETOWN 2018-02-15T17:30:53.000Z
## 23472 -77.07340 <NA> 2018-08-17T03:36:29.000Z
##
##          END_DATE OBJECTID  OCTO_RECOR  TRACT BLKGRP
## 1826          <NA> 259183436 18064713-01 000100      4
## 11452 2018-08-14T00:19:42.000Z 259193062 18134729-01 000100      4
## 26674 2018-03-14T03:49:12.000Z 259240847 18042061-01 000100      4
## 30745 2018-02-16T02:30:15.000Z 259245997 18026424-01 000100      4
## 30750 2018-02-16T08:15:53.000Z 259246002 18026435-01 000100      4
## 23472 2018-08-17T03:46:04.000Z 259224604 18136654-01 000201      1
##
##          geometry
## 1826 POINT (-77.06196 38.90259)
## 11452 POINT (-77.06068 38.9052)
## 26674 POINT (-77.06021 38.90684)
## 30745 POINT (-77.05759 38.90509)
## 30750 POINT (-77.05759 38.90509)
## 23472 POINT (-77.0734 38.91258)
```

```
dim(cbg)
```

```
## [1] 1567  26
```

Now we have a crime incident level simple feature (cbg) with block group information (TRACT,BLKGRP). We still need to create a block group level dataframe. First, we make the simple feature into a dataframe by getting rid of all geometry information. Recall, we no longer want to use the point-specific nature of this file.

```
st_geometry(cbg) <- NULL
```

Next, we want to aggregate to the block group level. We return to the pair of functions – `group_by()` and `summarize()` – that we learned in previous classes. Refer back to the previous tutorials for explanations of these. We check the number of observations after we summarize the data. It should have no more observations than the number of block groups times two, since each block group and offense gets its own row. There may be fewer observations than the number of block groups, since some block groups have no crime.

```
cbg2 <- group_by(cbg, TRACT, BLKGRP, OFFENSE)
cbg3 <- summarize(.data = cbg2, tot_crimes = n())
dim(cbg3)
```

```
## [1] 478  4
```

For the purposes of today’s tutorial, it is more useful to have a wide dataframe than a long one. A long dataframe is useful if you want to show different categories within one chart, such as categories within a histogram. However, if you want to make charts of multiple variables, a wide dataframe is more handy. Also, later in this tutorial, we want to make quartiles of variables. This is much easier to program when the data are wide.

Just as we used `gather()` to make a wide dataframe long, we now use `spread()` to make a long dataframe wide. We need to name the input data (`data =`), the “key” (the variable that identifies how the dataframe is currently long; for us “OFFENSE”), and the value of the variable, which for us is total crimes (`tot_crimes`).

```
# make this a wide dataset so we can merge with map
cbg4 <- spread(data = cbg3, key = OFFENSE, value = tot_crimes)
head(cbg4)
```

```
## # A tibble: 6 x 4
## # Groups:   TRACT, BLKGRP [6]
##   TRACT BLKGRP BURGLARY HOMICIDE
##   <fct> <fct>     <int>   <int>
## 1 000100 2             1       NA
```

```
## 2 000100 3          1      NA
## 3 000100 4          5      NA
## 4 000201 1          3      NA
## 5 000202 1          2      NA
## 6 000202 2          2      NA
```

Now that we have in hand a non-spatial block group file – non-spatial in the sense that it has plenty of information about each block group, but contains no geographic covariates – we need to add these data to the spatial block group file. Check the final file. How many observations should it have? How do you know? Answer these questions in the homework.

```
# merge each back into the block groups file so we have all block groups with shapes
bg2010.c2 <- merge(x = bg2010.small, y = cbg4, by = c("TRACT", "BLKGRP"), all = TRUE)
dim(bg2010.c2)
```

```
## [1] 450  5
```

C. Make quartiles for the map

With these data in hand, the one piece of information missing before making the graph is the number of quantiles – here we use quartiles – into which we want to divide the data. Recall that choropleth maps (usually) show discrete quantities, so we need tell R those categories.

To find quantiles, we use a R `dplyr` function called `ntile()`, which asks you for the input dataframe and variable, and then the number of groups into which you'd like to divide the data in roughly equal numbers.

You can check that this function is doing what you think it should be with the `table()` command. Each quartile should have roughly the same number of observations.

```
bg2010.c2$burg.quartile <- ntile(bg2010.c2$BURGLARY, 4)
table(bg2010.c2$burg.quartile)
```

```
##
##  1  2  3  4
## 96 95 95 95
```

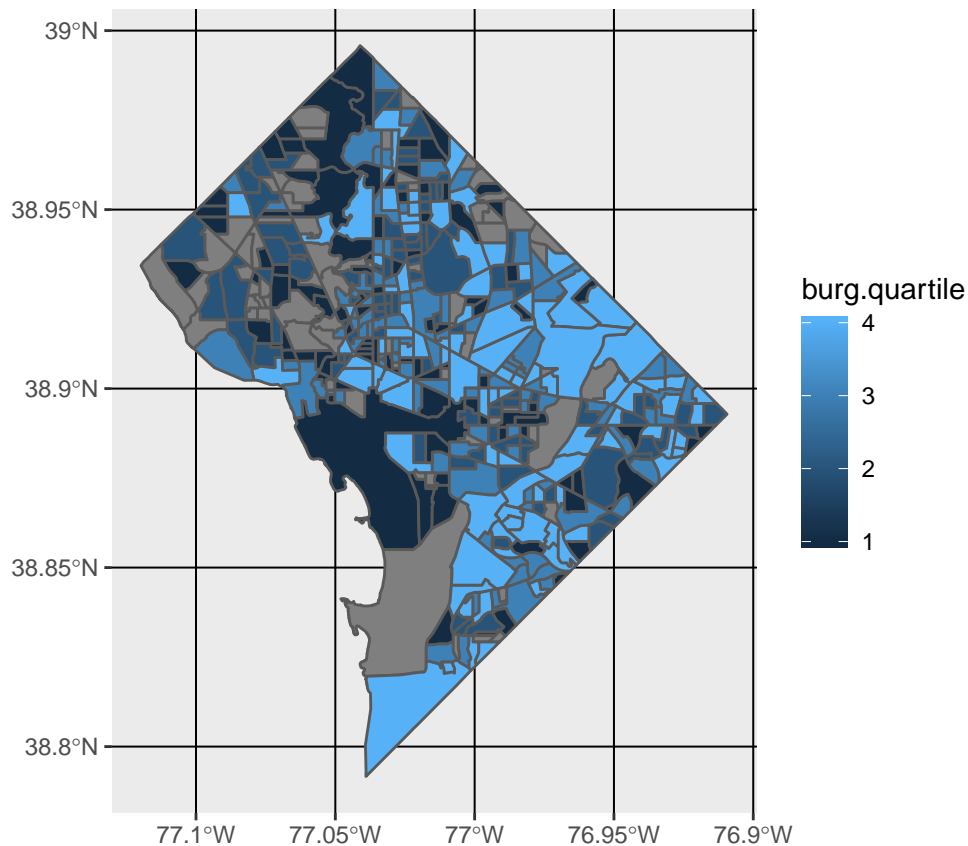
D. Choropleth maps

You're now ready to make a choropleth map. The basic idea is to use `geom_sf()` (as we did last mapping class to make maps) and `fill=` to plot any variable of interest.

D.1. Burglaries

We begin by plotting quartiles of burglaries (`burg.quartile`).

```
burg <- ggplot() +  
  geom_sf(data = bg2010.c2, aes(fill = burg.quartile))  
burg
```



This is a good first step. But this is far from the last step, because there are a number of things we should quite dislike about it

- legend has a graduated color scale, but the data are categorical
- low numbers are darker than high ones – attention should be directed to the darkest color, or the higher crime areas
- Some block groups have no crime and are in grey
- The plot background is useless and distracting

Let's fix these and see how things look. First, we make block groups with `NA` values zero. This is not always – in fact it is very rarely – a good idea. In many cases `NA` means that we don't know what the value is. However, in this particular case, we know that block groups with `NA` don't have crime, and that their true number of burglaries is really zero. Again, use `table()` to see if all observations now have a value.

```

# lets fix
# make NA values zero
bg2010.c2$BURGLARY <- ifelse(is.na(bg2010.c2$BURGLARY) == TRUE,0,bg2010.c2$BURGLARY)
table(bg2010.c2$burg.quartile,bg2010.c2$BURGLARY)

```

```

##
##      0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 16 17 20 23
##  1  0 92  4  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  2  0  0 79 16  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  3  0  0  0 38 48  9  0  0  0  0  0  0  0  0  0  0  0  0
##  4  0  0  0  0  0 20 28 14  6  5  7  1  2  2  5  2  1  1  1

```

The `table()` output shows that all one-burglary block groups are in the first quartile, along with a few two-burglary block groups. Two- and three-burglary block groups are in the second quartile, and the to quartile contains block groups with 5 to 23 burglaries.

To address the problem of the graduated color legend, let's create a color scale. I got these hex colors from (www.colorbrewer2.org)[<http://www.colorbrewer2.org/>]; we discuss this website in greater detail in the lecture. I create a vector (a list of values) that I can later use in the chart. Recall the `c()` notation for making a column.

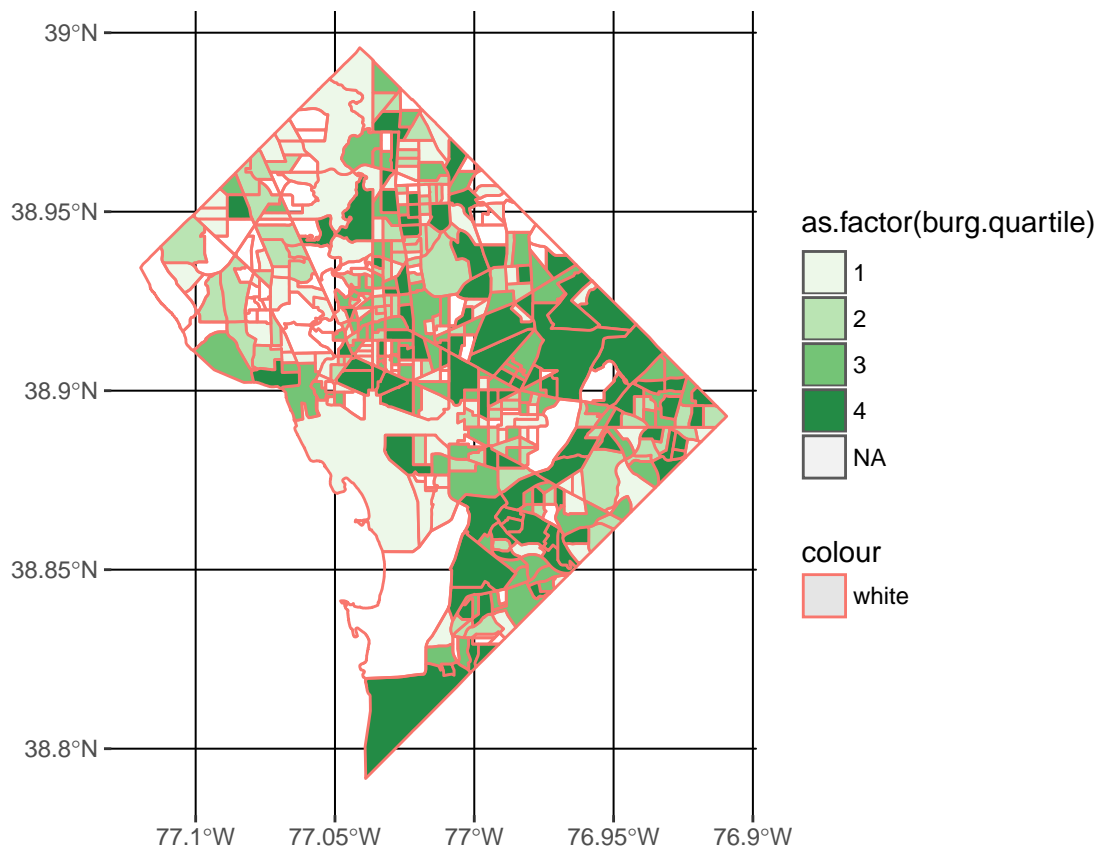
```

# make color ramp
# see
four.colors <- c("#edf8e9", "#bae4b3", "#74c476", "#238b45")

```

Now we try to make a graph with all of these fixes: without the NA values, without the plot background, and with better colors. For the latter, we use `scale_fill_manual()`, which will put in the colors in the order of the quartiles. If you don't like the order, you can either change the factor levels, or change the order of the colors. Note that we set both the panel and the plot background to be transparent.

```
# make graph with categorical legend
## still need to get rid of NAs and make borders white
burg <- ggplot() +
  geom_sf(data = bg2010.c2,
          aes(fill = as.factor(burg.quartile), color = "white")) +
  scale_fill_manual(values = four.colors) +
  theme(panel.background = element_rect(fill = "transparent", colour = NA),
        plot.background = element_rect(fill = "transparent", colour = NA))
burg
```



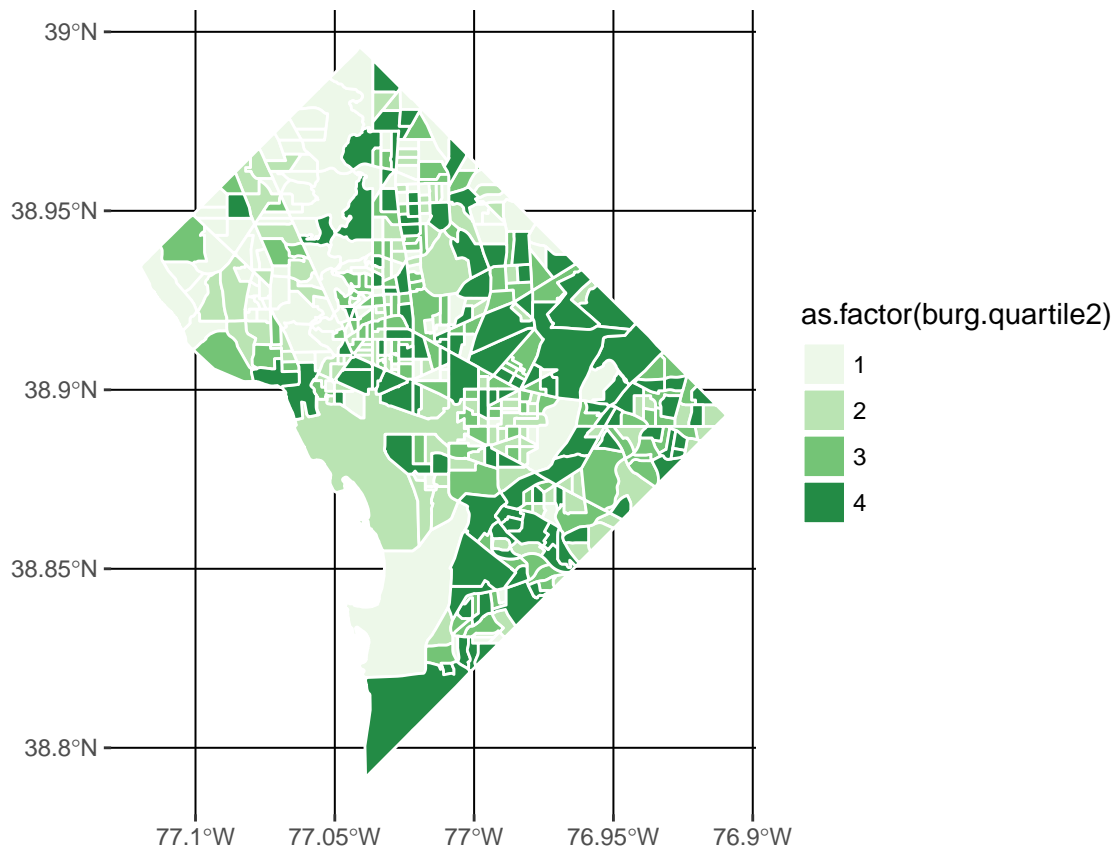
Some things are fixed in this revision, but we still have NAs. Why? Recall that we defined the quartiles before we set the NAs to zero – and the above graph is plotting those original quartiles. So the first step is to define new quartiles.

```
## for first problem: b/c quartiles were defined before we fixed the NAs to zeros
## lets try a second round of making quartiles
bg2010.c2$burg.quartile2 <- ntile(bg2010.c2$BURGLARY, 4)
table(bg2010.c2$burg.quartile2)
```

```
##
##  1  2  3  4
## 113 112 113 112
```


We also have an odd red border around the blockgroups. Looking carefully, this is because the `color=` command was *inside* the `aes()` command – where it should not have been. Fixing both of these issues, we get

```
# another attempt  
# the -white- was inside the aes call, which it should not have been  
# using the new quartiles solves the previous problem  
burg2 <- ggplot() +  
  geom_sf(data = bg2010.c2,  
          aes(fill = as.factor(burg.quartile2)), color = "white") +  
  scale_fill_manual(values = four.colors) +  
  theme(panel.background = element_rect(fill = "transparent", colour = NA),  
        plot.background = element_rect(fill = "transparent", colour = NA))  
burg2
```



This is not perfect, but it is good enough for now.

D.2. Homicides

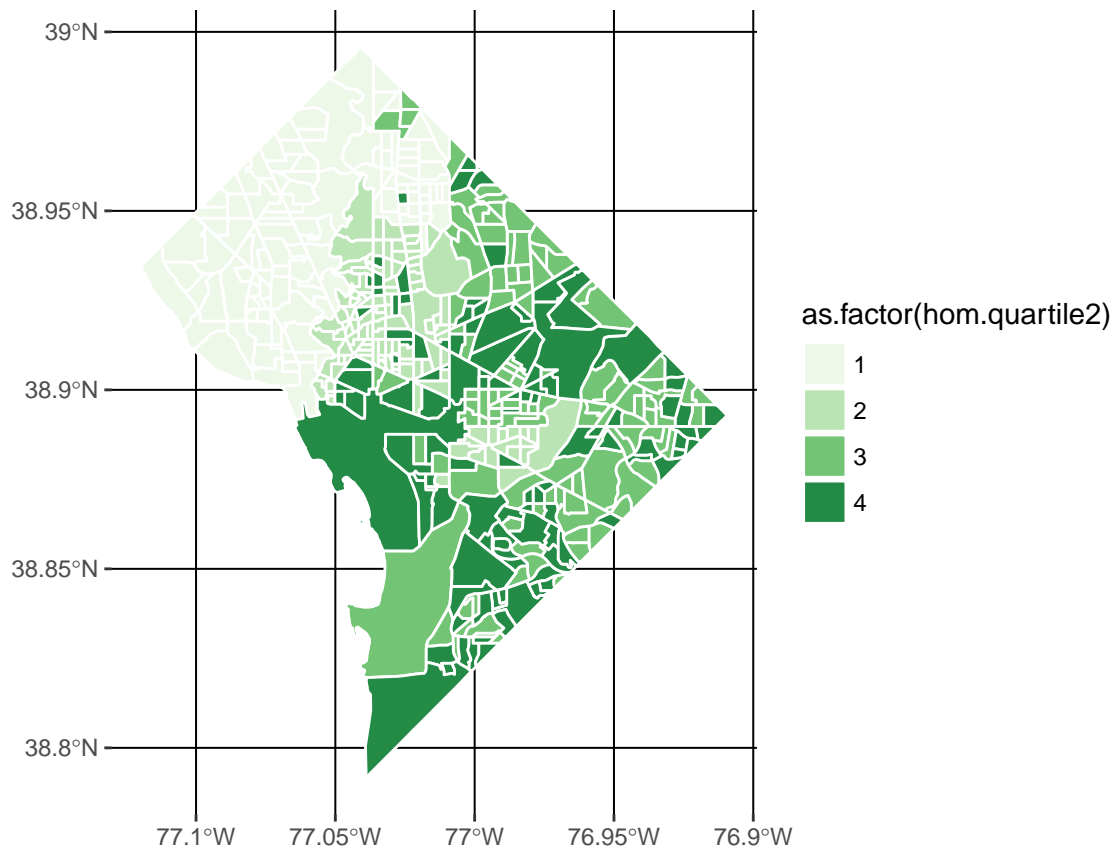
Let's repeat this exercise for homicides, of which there are (thankfully) many fewer than burglaries. First we set NAs equal to zero and then calculate quartiles (skipping the previous mistake!).

```
## look at homicide plot
bg2010.c2$HOMICIDE <- ifelse(is.na(bg2010.c2$HOMICIDE) == TRUE, 0, bg2010.c2$HOMICIDE)
bg2010.c2$hom.quartile2 <- ntile(bg2010.c2$HOMICIDE, 4)
table(bg2010.c2$hom.quartile2)
```

```
##
##  1  2  3  4
## 113 112 113 112
```

Plotting the same graph as before for burglaries, it looks like this

```
# another attempt
homc <- ggplot() +
  geom_sf(data = bg2010.c2,
          aes(fill = as.factor(hom.quartile2)), color = "white") +
  scale_fill_manual(values = four.colors) +
  theme(panel.background = element_rect(fill = "transparent", colour = NA),
        plot.background = element_rect(fill = "transparent", colour = NA))
homc
```



This graph is very misleading. take a look at the distribution of homicides and explain why it is misleading. Make another graph that is not as misleading.

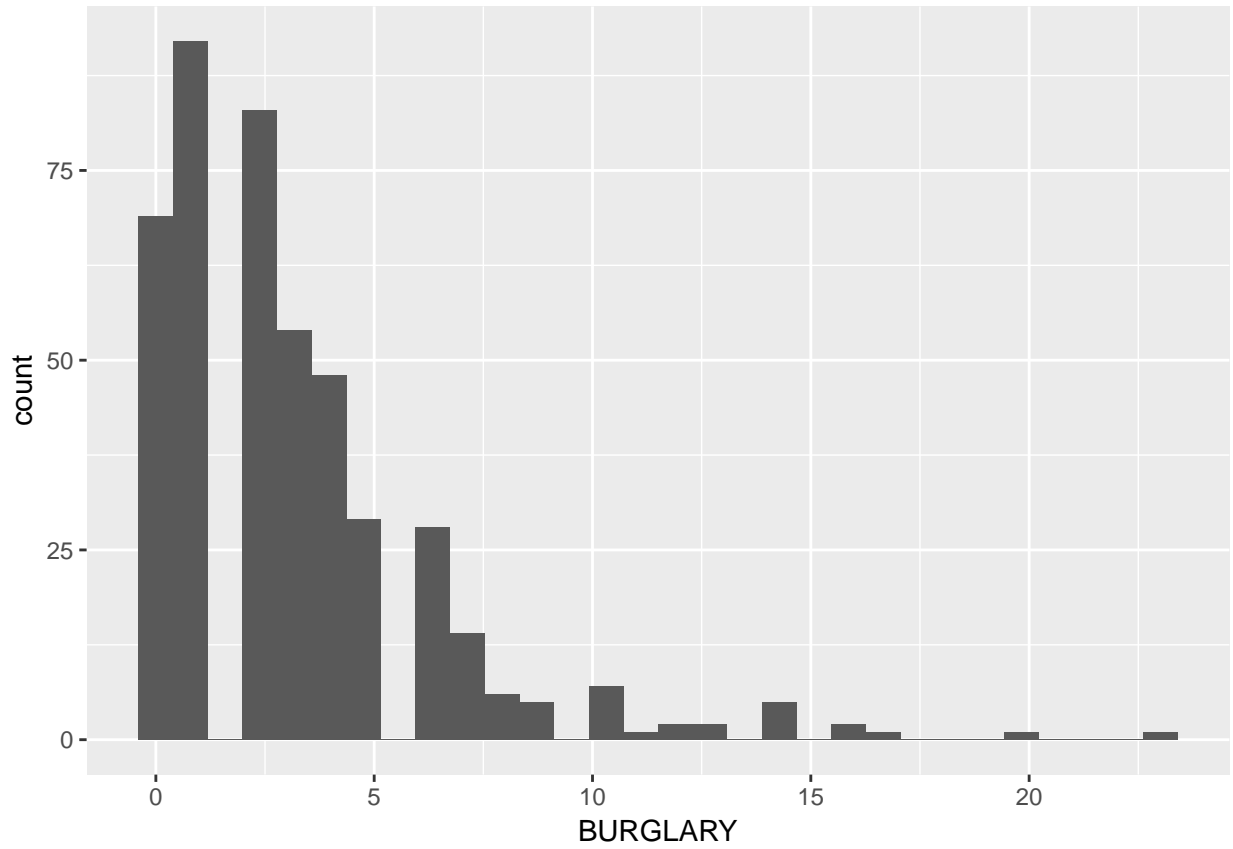
E. Histogram Legend

I have come to believe that all choropleth maps should be accompanied by a histogram legend. This is because, as we discuss in lecture, the histogram legend ameliorates part of the worst aspects of a choropleth map, namely the unconscious equation of size and value. It also helps us see the full distribution of the variable in a way that the map cannot (see this [great example](#)).

Below, we make a histogram showing the distribution of the number of burglaries, where the bars are colored the quartile in which the observation falls.

```
lego <- ggplot() +  
  geom_histogram(data = bg2010.c2,  
    aes(x = BURGLARY, fill = burg.quartile2))  
lego
```

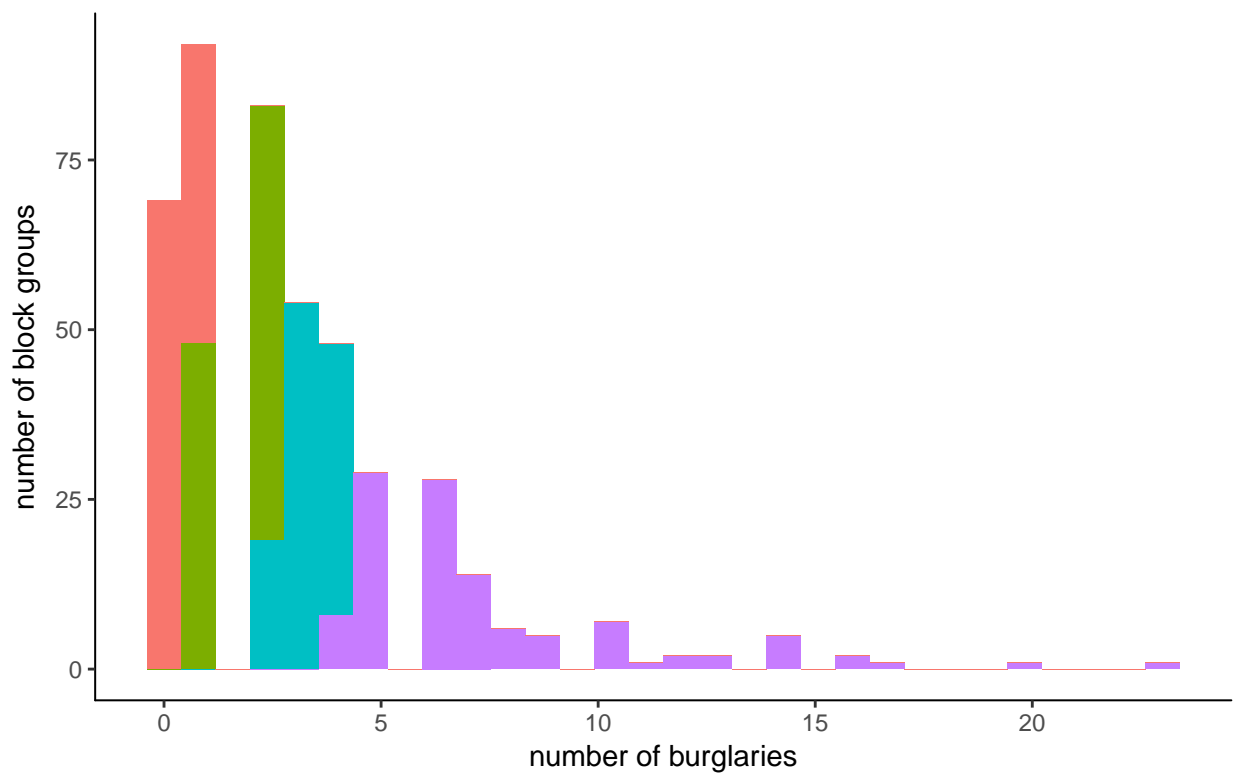
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



But all the bars are grey! This is because we tried to fill with a numeric variable. Let's try again with a factor. Let's also add axis labels, get rid of the background and get rid of the legend.

```
# note that this wont work with the fill as a numeric thing
lego <- ggplot() +
  geom_histogram(data = bg2010.c2,
    aes(x = BURGLARY, fill = as.factor(burg.quartile2))) +
  theme(legend.position = "none",
    axis.line.x = element_line(color="black"),
    axis.line.y = element_line(color="black"),
    panel.background = element_rect(fill = "transparent",colour = NA),
    plot.background = element_rect(fill = "transparent",colour = NA)) +
  labs(title = "", subtitle = "",
    x = "number of burglaries",
    y = "number of block groups")
lego
```

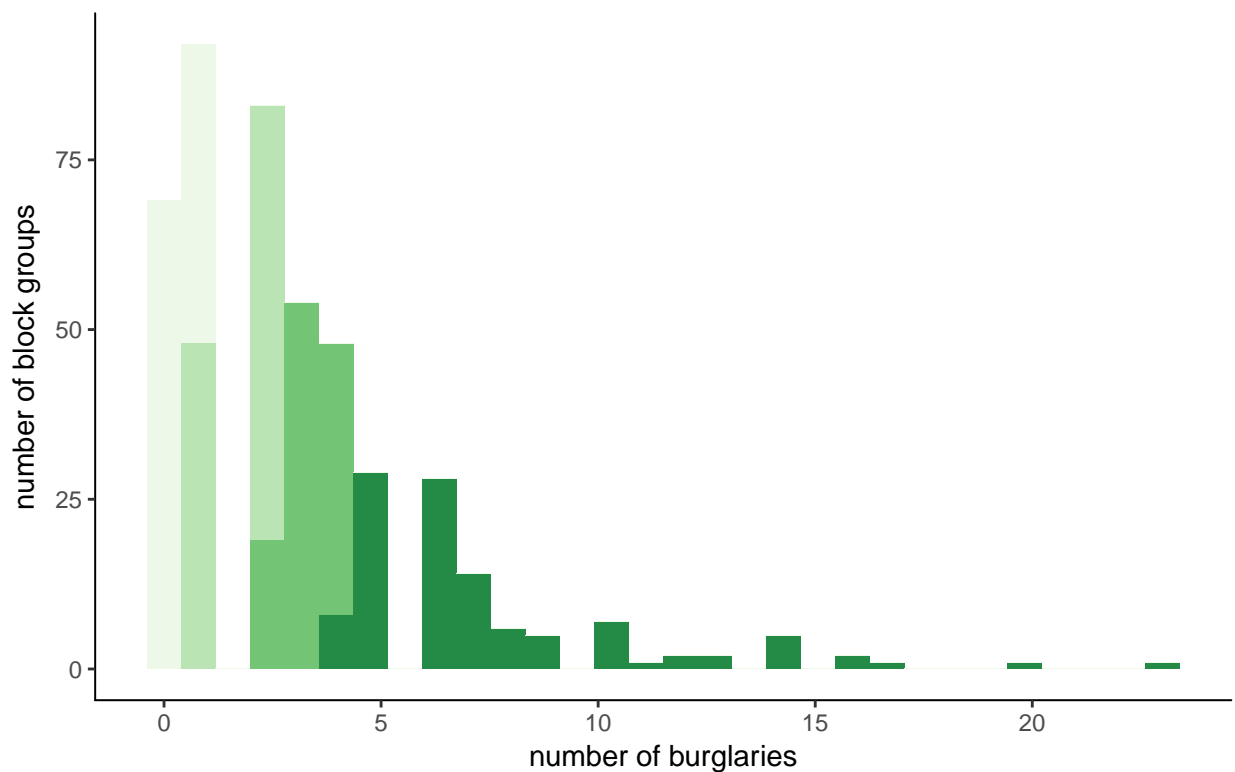
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



These are colors... but the legend only works if we use the *same* colors as the choropleth. I do this below by adding the `scale_fill_manual()` command from the choropleth maps.

```
# note that this wont work with the fill as a numeric thing
lego <- ggplot() +
  geom_histogram(data = bg2010.c2,
    aes(x = BURGLARY, fill = as.factor(burg.quartile2))) +
  scale_fill_manual(values = four.colors) +
  theme(legend.position = "none",
    axis.line.x = element_line(color="black"),
    axis.line.y = element_line(color="black"),
    panel.background = element_rect(fill = "transparent",colour = NA),
    plot.background = element_rect(fill = "transparent",colour = NA)) +
  labs(title = "", subtitle = "",
    x = "number of burglaries",
    y = "number of block groups")
lego
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



F. Dot Density Map

A dot density map is an alternative way of showing spatial patterns on a map. Here is a [very nice example](#) on which I based this section.

For this section, I needed to install a new package: `lwgeom`:

```
install.packages("lwgeom", dependencies = TRUE)
```

F.1. housing density

We begin with a density plot for housing units, with the intent of showing housing unit density.

To start, let's get a sense of what the housing unit variable looks like by block group, using `summary` and `dim`.

```
summary(bg2010$H0010001)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.0   413.2   583.5   659.4   843.8   3134.0
```

```
dim(bg2010)
```

```
## [1] 450 55
```

We learn that there are 450 block groups in DC, and the median block group has XXX houses. We probably cannot make one dot per house, since everything would look covered by (but you can check if you want!).

So let's rescale, so that we observe in units of 10 houses. We'll round up if there are any houses at all (alternatively, you could round based on 5). And, of course, we check the output.

```
# let's make each dot 10 houses, and then round up to 10 if < 10
bg2010$hby10 <- bg2010$H0010001/10
bg2010$hby10 <- ifelse(bg2010$hby10 < 10, 1, bg2010$hby10)
summary(bg2010$hby10)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00   41.33   58.35   65.94   84.38   313.40
```

The next step is a set of lines of code that get the points we need to put on the map. In order of appearance, here's what they do. The `st_sample()` command creates a matrix (`hdots`) of points from the simple feature file `bg2010`. It makes one point for each value of the variable equal to `size` (here the number of housing units divided by 10), and points are allocated randomly within a block group (or whatever polygon you've told R in `x=`).

This command generates *a lot* of error messages. I couldn't figure out how to suppress them in time for this tutorial, unfortunately. If you figure it out, let me know! Basically, they are saying that you are taking geographic coordinates and assuming the earth is flat for purposes of this sampling. Because our area is so small, that is perfectly fine.

But R does not yet know that the things in `hdots` are points, so we use `st_cast()` to "cast" them into points (`to = "POINT"`). Because we want to plot these points as a dataframe, not a simple feature (I'm not entirely sure why this is preferred, but it is what this author does, and I'm sure there must be a good reason!) we use `st_coordinates()` to get the values of the dots out the simple feature into a matrix.

But `ggplot` can't plot matrices, so we still need to create a dataframe using `as.data.frame()`, and then give names to the columns with `setNames`.

```
# get the points to plot
hdots <- st_sample(x = bg2010, size = bg2010$hby10, type = "random")
```

```
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
```

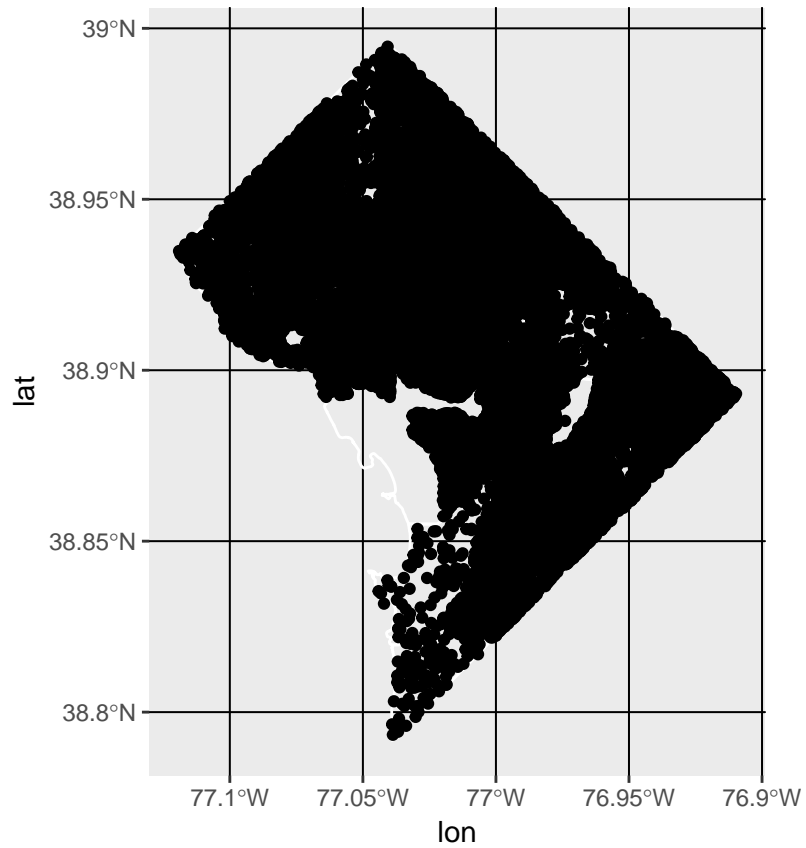


```
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
## although coordinates are longitude/latitude, st_intersects assumes that they are planar
```

```
hdots <- st_cast(x = hdots, to = "POINT")
hdots.mat <- st_coordinates(hdots)
hdats.df <- as.data.frame(hdots.mat)
hdats.df <- setNames(object = hdats.df, nm = c("lon", "lat"))
```


Now that we've put all these points together, let's plot, using our old tool of `geom_sf()` for the block group map and `geom_point()` for the points.

```
# now plot it with the DC bg map behind it  
h.map <-  
  ggplot() +  
    geom_sf(data = bg2010, fill = "transparent", color = "white") +  
    geom_point(data = hdats.df, aes(x=lon, y = lat))  
h.map
```

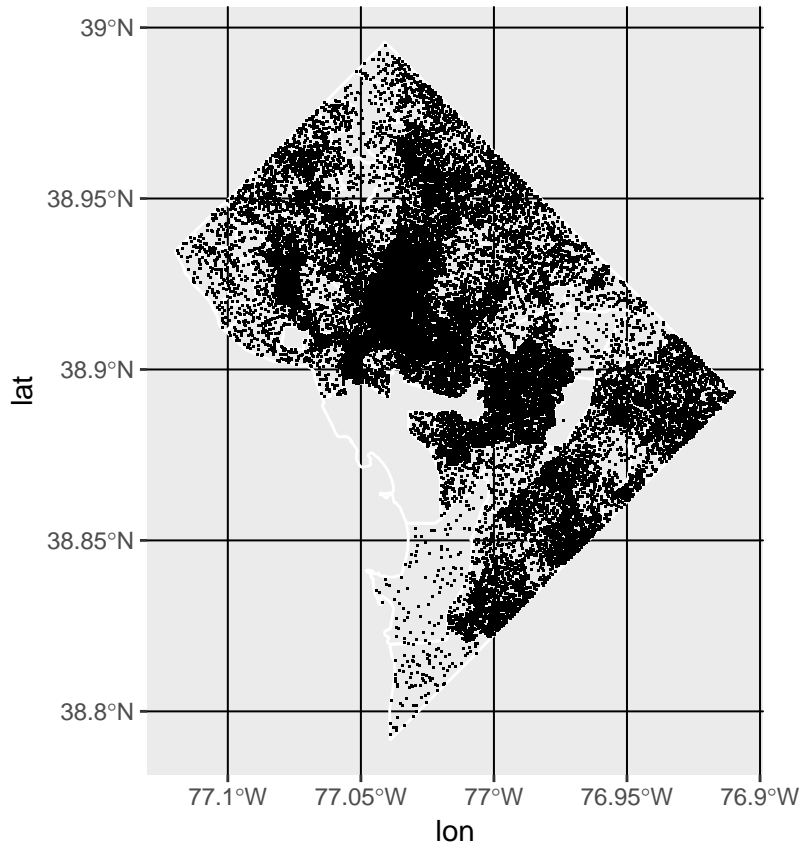


You can make this map better on your screen and see if it looks more legible. Alternatively, you can make small points using `shape = "."`.

```

# now plot it with the DC bg map behind it
h.map <-
  ggplot() +
    geom_sf(data = bg2010, fill = "transparent", color = "white") +
    geom_point(data = hdata.df, aes(x=lon, y = lat), shape = ".")
h.map

```



F.2. Dot density with two categories of things

This is just fine for one value, but I think the benefit of this method is when we want to show more than one value. In the previous graph, we could have equally well used a choropleth to make categories of housing units density (housing units divided by land area).

So now we'll make a map that shows occupied and vacant housing units. We'll also use the function skills we learned last class.

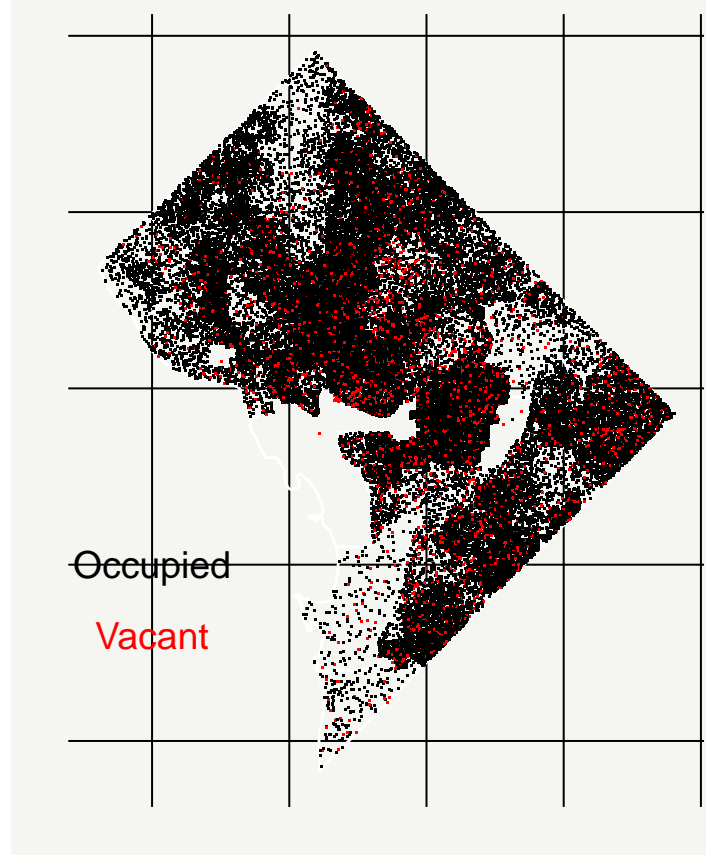
My function has three inputs: the variable we want to make into points (`hvar`), the scaling we want to use (by 5 units? 10 units? here it is by `scaler`), and the name of the variable (`namer`; this was useful for de-bugging, as well as for creating a ID variable).

This code is very similar to the previous, with two exceptions. Instead of scaling by 10, I'm using the function to scale by 5. You can try alternatives – which is easy, once you've built this into the function.

The second exception is that I add a column to the `hdata.df` dataframe with the name of the type of unit: occupied or vacant. So I use this function twice, once for occupied units and once for vacant ones, each time returning a dataframe.

And finally, we use `annotate()` to put the relevant legend info straight onto the map. This is much easier to understand than the legend that R auto-generates (get rid of `legend.position = "none"` to see for yourself).

```
# now plot it with the DC bg map behind it
oh.map <-
  ggplot() +
  geom_sf(data = bg2010, fill = "transparent", color = "white") +
  geom_point(data = hhdats.df, aes(x=lon, y=lat, color = htype), shape = ".") +
  scale_color_manual(values = c("black","red")) +
  labs(x="",y="") +
  theme(plot.background = element_rect(fill = "#f5f5f2", color = NA),
        panel.background = element_rect(fill = "#f5f5f2", color = NA),
        panel.grid = element_blank(),
        legend.background = element_rect(fill = "#f5f5f2", color = NA),
        legend.position = "none",
        axis.ticks = element_blank(),
        axis.text = element_blank()) +
  annotate(geom = "text",
         label = "Occupied",
         x = -77.1, y = 38.85,
         color = "black", size = 5) +
  annotate(geom = "text",
         x = -77.1, y = 38.83,
         label = "Vacant", color = "red",
         size = 5)
oh.map
```



G. Homework

1. Why is the homicide choropleth misleading? Fix it! (There are many ways to fix it – do anything you think is appropriate.)
2. Make choropleths for both homicides and burglaries in rates (per residential population). Recall that population is in bg2010 when you load it.
3. Pull in other block group data (the data we used earlier this semester is fine) and make a choropleth map with a histogram legend. Use terciles of the distribution.
4. Make a dot density map with the data of your choice and at least two groups.