

Creating informative maps

Based on Section 3.2.3 from Modern Data Science with R

```
# Initial packages required (we'll be adding more)
library(tidyverse)
library(mdsr)      # package associated with our MDSR book
```

Opening example

Here is a simple choropleth map example from MDSR

```
# CIACountries is a 236 x 8 data set with information on each country
# taken from the CIA factbook - gdp, education, internet use, etc.
head(CIACountries)
```

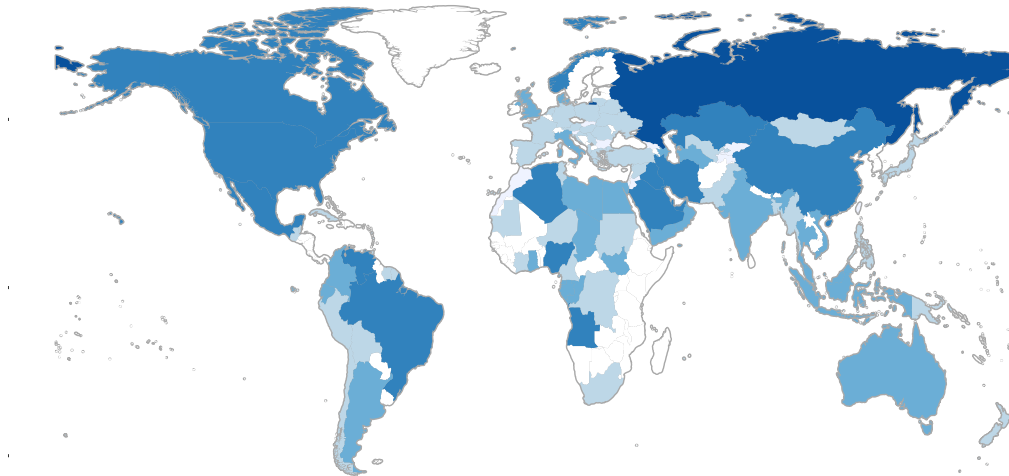
	country	pop	area	oil_prod	gdp	educ	roadways	net_users
1	Afghanistan	32564342	652230	0	1900	NA	0.06462444	>5%
2	Albania	3029278	28748	20510	11900	3.3	0.62613051	>35%
3	Algeria	39542166	2381741	1420000	14500	4.3	0.04771929	>15%
4	American Samoa	54343	199	0	13000	NA	1.21105528	<NA>
5	Andorra	85580	468	NA	37200	NA	0.68376068	>60%
6	Angola	19625353	1246700	1742000	7300	3.5	0.04125211	>15%

```
CIACountries |>
  select(country, oil_prod) |>
  mutate(oil_prod_disc = cut(oil_prod,
                             breaks = c(0, 1e3, 1e5, 1e6, 1e7, 1e8),
                             labels = c(">1000", ">10,000", ">100,000", ">1 million", ">10
# we won't use mWorldMap often, but it's a good quick illustration
mosaic::mWorldMap(key = "country") +
```

```
geom_polygon(aes(fill = oil_prod_disc)) +
scale_fill_brewer("Oil Prod. (bbl/day)", na.value = "white") +
theme(legend.position = "top")
```

Oil Prod. (bbl/day)

>1000	>100,000	>10 million
>10,000	>1 million	NA



Choropleth Maps

When you have specific regions (e.g. countries, states, counties, census tracts,...) and a value associated with each region.

A choropleth map will color the entire region according to the value. For example, let's consider state vaccination data from March 2021.

```
vaccine_data <- read_csv("Data/vacc_Mar21.csv")

vacc_mar13 <- vaccine_data |>
  filter(Date == "2021-03-13") |>
  select(State, Date, people_vaccinated_per100, share_doses_used, Governor)

vacc_mar13
```

A tibble: 50 x 5

	State	Date	people_vaccinated_per100	share_doses_used	Governor
	<chr>	<date>	<dbl>	<dbl>	<chr>
1	Alabama	2021-03-13	17.2	0.671	R
2	Alaska	2021-03-13	27.0	0.686	R
3	Arizona	2021-03-13	21.5	0.821	R
4	Arkansas	2021-03-13	19.2	0.705	R
5	California	2021-03-13	20.3	0.726	D
6	Colorado	2021-03-13	20.8	0.801	D
7	Connecticut	2021-03-13	26.2	0.851	D
8	Delaware	2021-03-13	20.2	0.753	D
9	Florida	2021-03-13	20.1	0.766	R
10	Georgia	2021-03-13	15.2	0.674	R

i 40 more rows

The tricky part of choropleth maps is getting the shapes (polygons) that make up the regions. This is really a pretty complex set of lines for R to draw!

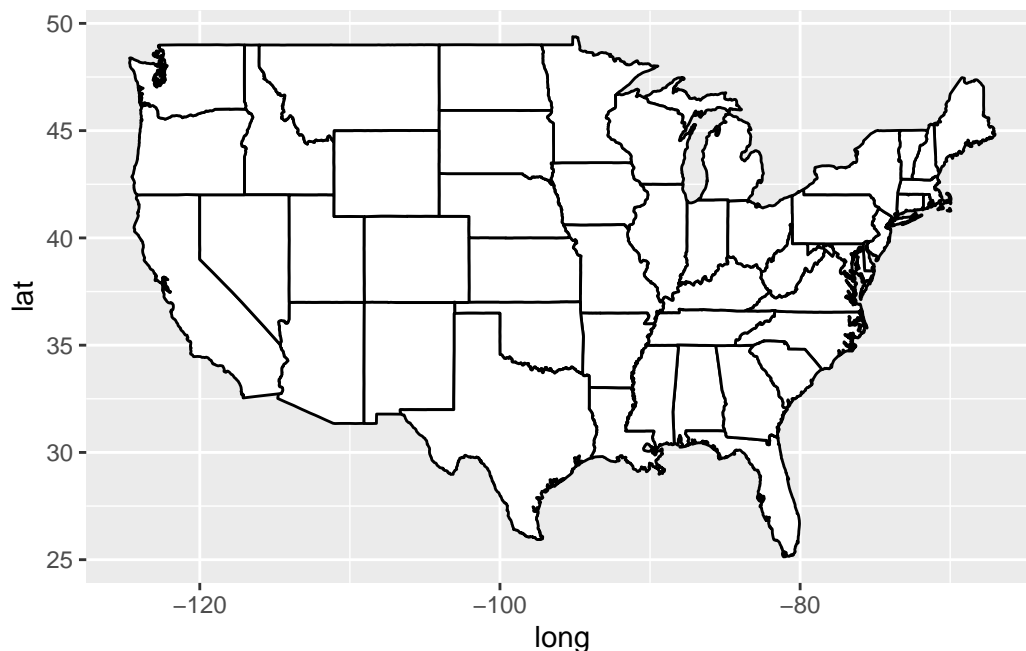
Luckily, some maps are already created in R in the maps package.

```
library(maps)
us_states <- map_data("state")
head(us_states)
```

	long	lat	group	order	region	subregion
1	-87.46201	30.38968	1	1	alabama	<NA>
2	-87.48493	30.37249	1	2	alabama	<NA>
3	-87.52503	30.37249	1	3	alabama	<NA>
4	-87.53076	30.33239	1	4	alabama	<NA>
5	-87.57087	30.32665	1	5	alabama	<NA>
6	-87.58806	30.32665	1	6	alabama	<NA>

Note that points in the same "group" are connected with a line

```
us_states |>
  ggplot(mapping = aes(x = long, y = lat,
                       group = group)) +
  geom_polygon(fill = "white", color = "black")
```



Other maps provided by the `maps` package include US counties, France, Italy, New Zealand, and two different views of the world. If you want maps of other countries or regions, you can often find them online.

Sometimes maps may be provided as shapefiles. To use these, you'll first need to read them into R and then turn them into tidy dataframes in order to use them with ggplot. See [here](#). More on shapefiles in Part 2.

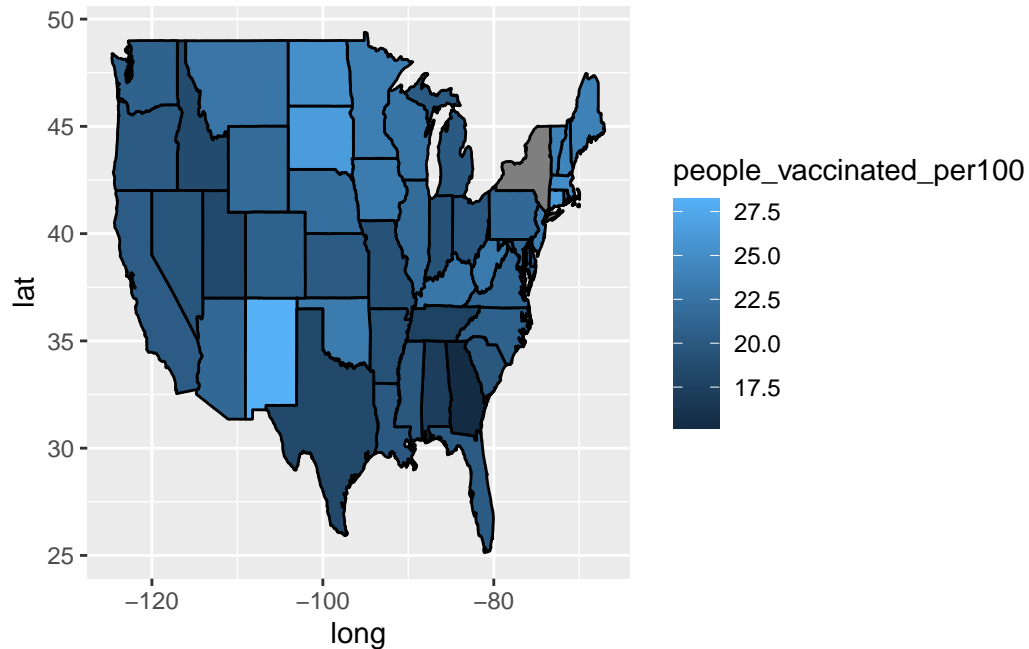
Where the really cool stuff happens is when we join our data to the `us_states` dataframe. Notice that the state name appears in the “region” column of `us_states`, and that the state name is in all small letters. In `vacc_mar13`, the state name appears in the `State` column and is in title case. Thus, we have to be very careful when we join the state vaccine info to the state geography data.

Run this line by line to see what it does:

```
vacc_mar13 <- vacc_mar13 |>
  mutate(State = str_to_lower(State))

vacc_mar13 |>
  right_join(us_states, by = c("State" = "region")) |>
  rename(region = State) |>
  ggplot(mapping = aes(x = long, y = lat,
```

```
group = group)) +  
geom_polygon(aes(fill = people_vaccinated_per100), color = "black")
```



oops, New York appears to be a problem.

```
vacc_mar13 |>  
anti_join(us_states, by = c("State" = "region"))
```

A tibble: 3 x 5

	State	Date	people_vaccinated_per100	share_doses_used	Governor
	<chr>	<date>	<dbl>	<dbl>	<chr>
1	alaska	2021-03-13	27.0	0.686	R
2	hawaii	2021-03-13	22.8	0.759	D
3	new york state	2021-03-13	21.7	0.764	D

```
us_states |>  
anti_join(vacc_mar13, by = c("region" = "State")) |>  
count(region)
```

```
region    n
```

```
1 district of columbia 10
2                new york 495
```

[Pause to ponder:] What did we learn by running `anti_join()` above?

Notice that the `us_states` map also includes only the contiguous 48 states. [This](#) gives an example of creating really beautiful map insets for Alaska and Hawaii.

```
vacc_mar13 <- vacc_mar13 |>
  mutate(State = str_replace(State, " state", ""))

vacc_mar13 |>
  anti_join(us_states, by = c("State" = "region"))
```

```
# A tibble: 2 x 5
  State Date      people_vaccinated_per100 share_doses_used Governor
<chr> <date>          <dbl>          <dbl> <chr>
1 alaska 2021-03-13      27.0          0.686 R
2 hawaii 2021-03-13      22.8          0.759 D
```

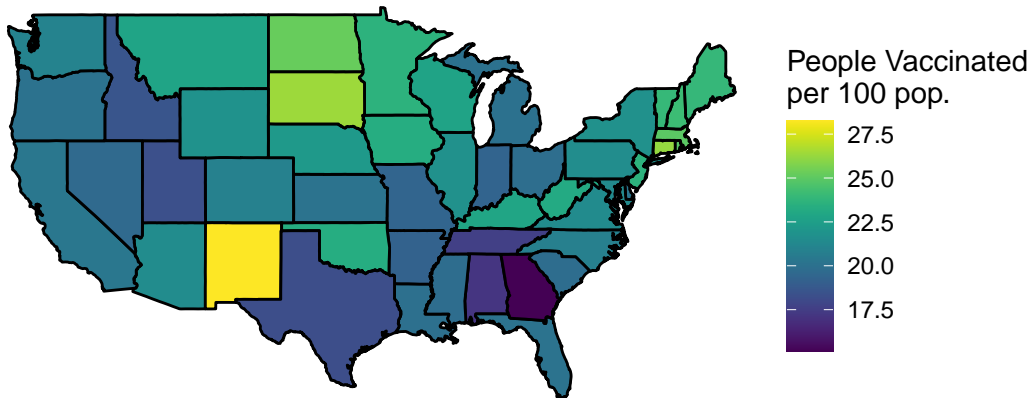
```
us_states |>
  anti_join(vacc_mar13, by = c("region" = "State")) %>%
  count(region)
```

```
      region n
1 district of columbia 10
```

Better.

```
library(viridis) # for color schemes
vacc_mar13 |>
  right_join(us_states, by = c("State" = "region")) |>
  rename(region = State) |>
  ggplot(mapping = aes(x = long, y = lat,
                      group = group)) +
  geom_polygon(aes(fill = people_vaccinated_per100), color = "black") +
  labs(fill = "People Vaccinated\nper 100 pop.") +
  # This scales the longitude and latitude so that the shapes look correct.
  coord_map() +
  # This theme can give you a really clean look!
```

```
theme_void() +
# you can change the fill scale for different color schemes.
scale_fill_viridis()
```



[Pause to ponder:] Use `autofill` to play with different themes and `scale_fills`.

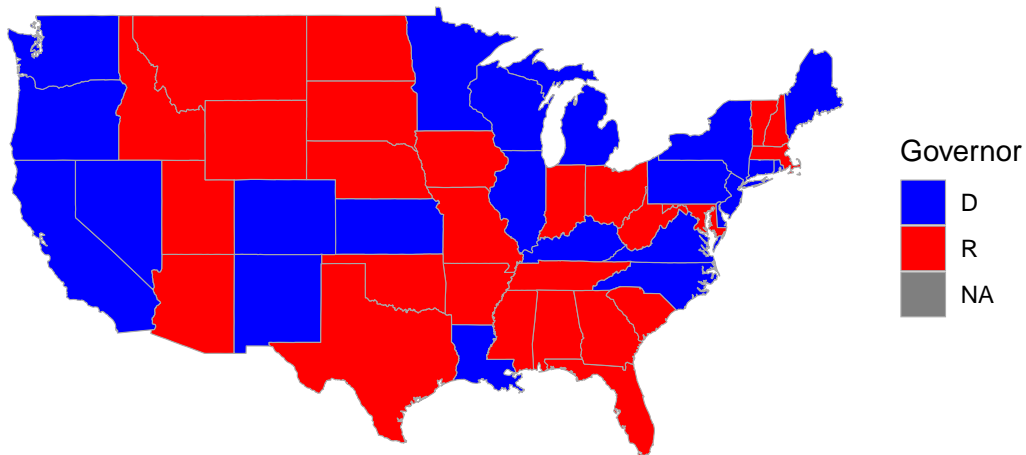
Note: Map projections are actually pretty complicated, especially if you're looking at large areas (e.g. world maps). It's impossible to preserve both shape and area when projecting a sphere onto a flat surface, so that's why you sometimes see [such different maps of the world](#)

There are a few different options in `coord_map()`. See the help menu, although this function is being phased out.

You can also use a categorical variable to color regions:

```
vacc_mar13 |>
  right_join(us_states, by = c("State" = "region")) |>
  rename(region = State) |>
  ggplot(mapping = aes(x = long, y = lat,
                       group = group)) +
  geom_polygon(aes(fill = Governor), color = "darkgrey", linewidth = 0.2) +
  labs(fill = "Governor") +
```

```
# This scales the longitude and latitude so that the shapes look correct.
coord_map() +
# This theme can give you a really clean look!
theme_void() +
# you can change the fill scale for different color schemes.
scale_fill_manual(values = c("blue", "red"))
```



Multiple maps!

[Pause to ponder:] are we bothered by the warning about many-to-many when you run the code below?

```
library(lubridate)
weekly_vacc <- vaccine_data |>
  mutate(State = str_to_lower(State)) |>
  mutate(State = str_replace(State, " state", ""),
         week = week(Date)) |>
  group_by(week, State) |>
  summarize(date = first(Date),
           mean_daily_vacc = mean(daily_vaccinated/est_population*1000)) |>
  right_join(us_states, by = c("State" = "region")) |>
```

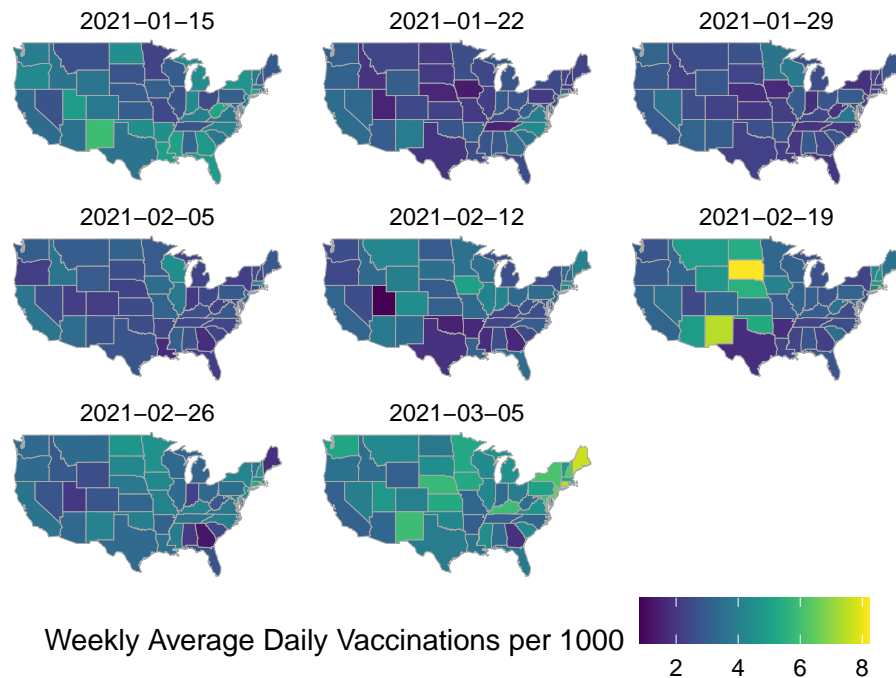


```

  rename(region = State)

weekly_vacc |>
  filter(week > 2, week < 11) |>
  ggplot(mapping = aes(x = long, y = lat,
                       group = group)) +
  geom_polygon(aes(fill = mean_daily_vacc), color = "darkgrey", size = 0.1) +
  labs(fill = "Weekly Average Daily Vaccinations per 1000") +
  coord_map() +
  theme_void() +
  scale_fill_viridis() +
  facet_wrap(~date) +
  theme(legend.position = "bottom")

```



Other cool state maps

statebin (square representation of states)

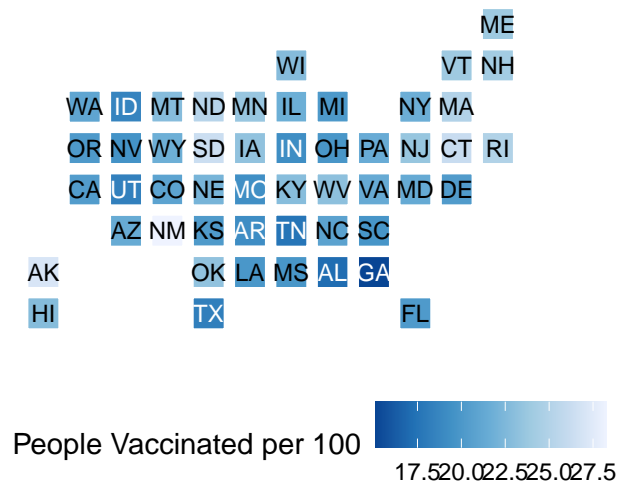
```

library(statebins) # may need to install

vacc_mar13 |>

```

```
mutate(State = str_to_title(State)) |>
statebins(state_col = "State",
          value_col = "people_vaccinated_per100") +
# one nice layout. You can customize with usual ggplot themes.
theme_statebins() +
labs(fill = "People Vaccinated per 100")
```



[Pause to ponder:] Why might one use a map like above instead of our previous choropleth maps?

I used [this example](#) to create the code above. The original graph is located [here](#).

Interactive map with leaflet

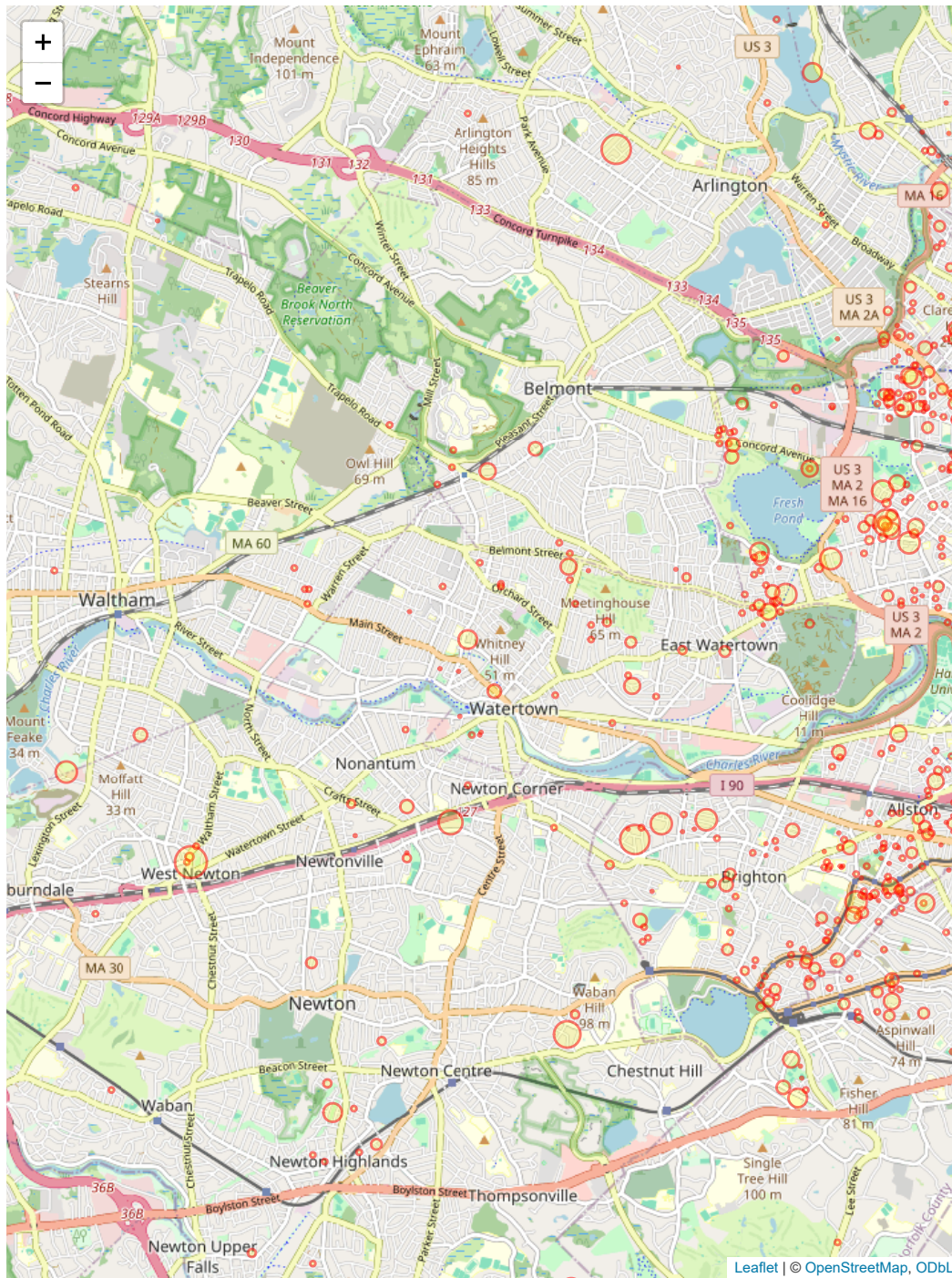
With leaflet, you can have “pop-up” messages when you hover over points, and have a zoom-in and zoom-out option. Leaflet is not part of the tidyverse, but the structure of its code is pretty similar.

Two main features: `addTiles()` = Add background map `setView()` = Set where the map should originally zoom to

```
# This part is for the pop-up messages.... some are weird or just "\n" for example, so thi
Encoding( x = airbnb.df$AboutListing ) <- "UTF-8"
airbnb.df$AboutListing <-
  iconv( x = airbnb.df$AboutListing
        , from = "UTF-8"
        , to = "UTF-8"
        , sub = "" )
head(airbnb.df)
```

```
# A tibble: 6 x 65
  ListingID Title      UserID baseurl Price AboutListing HostName MemberDate Lat
    <dbl> <chr>      <dbl> <chr>   <dbl> <chr>      <chr>      <chr>      <dbl>
1 281552 Harvard~ 1.47e6 https:~ 175 "\n"      Mary Ca~ December ~ 42.4
2 182613 Luxury ~ 8.76e5 https:~ 249 "Entire lar~ Max      July 2011 42.4
3 1587540 Cozy Ho~ 2.00e6 https:~ 225 "\n"      Finola   March 2012 42.4
4 469506 Luxury ~ 1.77e6 https:~ 140 "\n"      Rupal    February ~ 42.3
5 3937268 Boston ~ 2.53e6 https:~ 99 "I offer a ~ Natasha June 2012 42.3
6 3036349 Top flo~ 1.37e6 https:~ 89 "Two bedroo~ Carol    November ~ 42.3
# i 56 more variables: Long <dbl>, BookInstantly <chr>, Cancellation <chr>,
# PageCounter <dbl>, PageNumber <dbl>, A_AC <dbl>, A_Breakfast <dbl>,
# A_CableTV <dbl>, A_CarbonMonoxDetector <dbl>, A_Doorman <dbl>,
# A_Dryer <dbl>, A_TV <dbl>, A_Elevator <dbl>, A_Essentials <dbl>,
# A_Events <dbl>, A_FamilyFriendly <dbl>, A_FireExt <dbl>, A_Fireplace <dbl>,
# A_FirstAidKit <dbl>, A_Gym <dbl>, A_Heat <dbl>, A_HotTub <dbl>,
# A_Intercom <dbl>, A_Internet <dbl>, A_Kitchen <dbl>, A_Parking <dbl>, ...
```

```
# This part makes the map!
leaflet() |>
  addTiles() |>
  setView(lng = mean(airbnb.df$Long), lat = mean(airbnb.df$Lat),
        zoom = 13) |>
  addCircleMarkers(data = airbnb.df,
    lat = ~ Lat,
    lng = ~ Long,
    popup = ~ AboutListing,
    radius = ~ S_Accommodates,
    # These last options describe how the circles look
    weight = 2,
    color = "red",
    fillColor = "yellow")
```



On Your Own

The `states` dataset in the `poliscidata` package contains 135 variables on each of the 50 US states. See [here](#) for more detail.

Your task is to create two meaningful choropleth plots, one using a numeric variable and one using a categorical variable from `states`. Write a sentence or two describing what you can learn from each plot.

Here's some R code to get you going:

```
library(poliscidata) # may have to install first

# Be sure you know what the mutate statement below is doing!
state_data <- as_tibble(poliscidata::states) |>
  mutate(state_name = str_squish(str_to_lower(as.character(state)))) |>
  select(-state)
print(state_data, n = 5, width = Inf)
```

A tibble: 50 x 135

	abort_rank3 <fct>	abortion_rank12 <dbl>	adv_or_more <dbl>	ba_or_more <dbl>	cig_tax12 <dbl>	cig_tax12_3 <fct>
1	Less restr	35	9	26.6	2	HiTax
2	Mid	20	7.7	22	0.425	LoTax
3	More restr	4	6.1	18.9	1.15	MidTax
4	More restr	5	9.3	25.6	2	HiTax
5	Less restr	49	10.7	29.9	0.87	MidTax

	conserv_advantage <dbl>	conserv_public <dbl>	dem_advantage <dbl>	govt_worker <dbl>	gun_rank3 <fct>
1	21.3	43.1	-12.2	28	Less restr
2	36	44.7	-14.6	17.5	Mid
3	26.7	45.2	-1.4	17.6	Less restr
4	19.5	36	-3.5	15.5	Less restr
5	6.3	30.8	14.9	14.9	More restr

	gun_rank11 <dbl>	gun_scale11 <dbl>	hr_cons_rank11 <dbl>	hr_conserv11 <dbl>	hr_lib_rank11 <dbl>	hr_liberal11 <dbl>
1	50	0	200	55.7	228	44.3
2	17	14	152.	65.6	278.	34.4
3	39	4	132.	69.3	295.	30.7
4	50	0	156.	62.6	270.	37.4
5	1	81	274.	54.8	152.	81.0

	hs_or_more <dbl>	obama2012 <dbl>	obama_win12 <fct>	pop2000 <dbl>	pop2010 <dbl>	pop2010_hun_thou <dbl>
--	---------------------	--------------------	----------------------	------------------	------------------	---------------------------

1	91.4	40.8	No	626932	710231	7.10			
2	82.1	38.4	No	4447100	4779736	47.8			
3	82.4	36.9	No	2673400	2915918	29.2			
4	84.2	44.4	No	5130632	6392017	63.9			
5	80.6	60.2	Yes	33871648	37253956	373.			
	popchnng0010	popchnngpct	pot_policy	prochoice	prolife	relig_cath			
	<dbl>	<dbl>	<fct>	<dbl>	<dbl>	<dbl>			
1	83299	13.3	Medicinal	58	37	14.6			
2	332636	7.5	Decrim	36	54	6.6			
3	242518	9.1	Not legal	40	55	5.9			
4	1261385	24.6	Medicinal	56	39	27.3			
5	3382308	10	Medicinal / Decrim.	65	28	31.9			
	relig_prot	relig_high	relig_low	religiosity3	romney2012	smokers12	stateid		
	<dbl>	<dbl>	<dbl>	<fct>	<dbl>	<dbl>	<fct>		
1	50	31.3	39.5	Low	54.8	24	"AK"		
2	79.3	55.7	14.3	High	60.6	25	"AL"		
3	78.6	52.3	18.7	High	60.6	26	"AR"		
4	43.3	36.6	33.9	Mid	53.5	21	"AZ"		
5	37.8	34.5	36.6	Low	37.1	15	"CA"		
	to_0812	uninsured_pct	abort_rate05	abort_rate08	abortlaw3	abortlaw10	alcohol		
	<dbl>	<dbl>	<dbl>	<dbl>	<fct>	<dbl>	<dbl>		
1	-9.40	21.8	13.6	12	0-5 restr	5	3.02		
2	-2.90	18.8	11.9	12	6-8 restr	8	2.01		
3	-2.90	21.9	8.3	8.7	9-10 restr	9	1.83		
4	-3.10	20.5	16	15.2	6-8 restr	6	2.31		
5	-6.5	23.2	27.1	27.6	0-5 restr	4	2.33		
	attend_pct	battle04	blkleg	blkpct04	blkpct08	blkpct10	bush00	bush04	carfatal
	<dbl>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	22	No	2	3.6	4.3	4.7	58.6	61.1	17.4
2	52	No	25	26.4	26.4	26.8	56.5	62.5	24.9
3	50	No	11	15.8	15.8	16.1	51.3	54.3	25.6
4	29	No	1	3.5	4.2	5	51.0	54.8	20.3
5	33	No	5	6.8	6.7	7.2	41.7	44.4	12.1
	carfatal07	cig_tax	cig_tax_3	cigarettes	college	compct_m	cons_hr06	cons_hr09	
	<dbl>	<dbl>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	15.2	2	\$1.41-\$2.58	6.22	26.7	36.3	72	75	
2	25.9	0.425	\$0.07-\$0.64	9.41	21.1	40.7	77.7	72	
3	23.7	0.59	\$0.07-\$0.64	8.51	19.1	38.9	56.2	28.5	
4	17.6	2	\$1.41-\$2.58	2.4	24.3	33.3	69	49.5	
5	11.7	0.87	\$0.695-\$1.36	3.69	29.1	28.5	37.3	35.1	
	cook_index	cook_index3	defexpen	demhr11	dem_hr09	demnat06	dempct_m	demstate06	
	<dbl>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	-13.4	More Rep	3556	0	0	0	26.1	43.3	

2	-13.2	More Rep	1757	14.3	42.9	22.2	38.9	60.7
3	-8.8	More Rep	530	25	75	83.3	43.1	75.6
4	-6.1	Even	1771	28.6	62.5	20	31.9	44.4
5	7.4	More Dem	1106	64.2	64.2	63.6	41.3	60.8
	demstate09	demstate13	density	division	earmarks_pcap	evm	evo	evo2012
	<dbl>	<dbl>	<dbl>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>
1	46.7	36.7	1.2	Pacific	426.	3	0	0
2	57.9	35.7	94.4	E. South Cent	38.9	9	0	0
3	72.6	46.7	56	W. South Cent	26.7	6	0	0
4	41.1	41.1	56.3	Mountain	20.9	10	0	0
5	64.2	66.7	239.	Pacific	12.5	0	55	55
	evr2012	gay_policy	gay_policy2	gay_policy_con	gay_support	gay_support3		
	<dbl>	<fct>	<fct>	<fct>	<dbl>	<fct>		
1	3	Conservative	Conservative	No	56	Med		
2	9	Most conservative	Conservative	Yes	44	Low		
3	6	Most conservative	Conservative	Yes	44	Low		
4	11	Conservative	Conservative	No	58	Med		
5	0	Liberal	Liberal	No	64	High		
	gb_win00	gb_win04	gore00	gun_check	gun_dealer	gun_murder10	gun_rank_rev	
	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	Bush win	Bush win	27.7	12016.	139.	2.7	6	
2	Bush win	Bush win	41.6	9025.	47.4	2.8	30	
3	Bush win	Bush win	45.9	8443.	67.4	3.2	13	
4	Bush win	Bush win	44.7	5314.	45.4	3.6	13	
5	Gore win	Kerry win	53.4	3040.	21.6	3.4	48	
	gunlaw_rank	gunlaw_rank3_rev	gunlaw_scale	hispanic04	hispanic08	hispanic10		
	<dbl>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>		
1	43	Fewer restr	4	4.9	6.1	5.5		
2	19	Mid	15	2.2	2.9	3.9		
3	36	Fewer restr	6	4.4	5.6	6.4		
4	36	Fewer restr	6	28	30.1	29.6		
5	1	More restr	79	34.7	36.6	37.6		
	indpct_m	kerry04	libpct_m	mccain08	modpct_m	nader00	obama08	obama_win08
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<fct>	<dbl>
1	43.6	35.5	17.9	59.4	45.7	10.1	37.9	McCain win
2	30.0	36.8	16.8	60.3	42.5	1.10	38.7	McCain win
3	35.9	44.6	16.8	58.7	44.3	1.46	38.9	McCain win
4	29.7	44.4	19.2	53.4	47.4	2.98	44.9	McCain win
5	25.8	54.3	24.2	36.9	47.3	3.82	60.9	Obama win
	permit	pop_18_24	pop_18_24_10	prcapinc	region	relig_import	religiosity	
	<dbl>	<dbl>	<dbl>	<dbl>	<fct>	<dbl>	<dbl>	
1	NA	11.1	10.6	34454	West	NA	-177	
2	27.6	10.0	10.0	27795	South	58.5	-13	

3	21.1	10.1	9.74	25725	South	53.1	-23
4	46.2	9.61	9.90	28442	West	33.2	-140
5	52.8	9.95	10.5	35019	West	28.8	-147
	reppct_m	rtw	secularism	secularism3	seniority_sen2	south	to_0004 to_0408
	<dbl>	<fct>	<dbl>	<fct>	<fct>	<fct>	<dbl> <dbl>
1	30.3	No	177	Secular	Yes	Nonsouth	-12.4 -0.800
2	31.2	Yes	13	Religious	No	South	4.2 4.60
3	21.0	Yes	23	Religious	No	South	3.31 -0.200
4	38.3	Yes	140	Secular	No	Nonsouth	-2.32 1.90
5	32.9	No	147	Secular	No	Nonsouth	6.63 2.90
	trnout00	trnout04	unemploy	union04	union07	union10	urban vep00_turnout
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	67.2	54.8	7.5	20.1	23.8	22.3	65.6 68.1
2	50.6	54.8	5.8	9.7	9.5	10.9	55.4 51.6
3	46.9	50.2	5.9	4.8	5.4	4.2	52.5 47.9
4	44.6	42.3	5.1	6.3	8.8	6.5	88.2 45.6
5	54.6	61.2	6.2	16.5	16.7	17.2	94.4 55.7
	vep04_turnout	vep08_turnout	vep12_turnout	womleg_2007	womleg_2010	womleg_2011	
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	69.1	68.3	58.9	21.7	21.7	23.3	
2	57.2	61.8	58.9	12.9	12.9	13.6	
3	53.6	53.4	50.5	20.7	23	22.2	
4	54.1	56	52.9	34.4	31.1	34.4	
5	58.8	61.7	55.2	28.3	27.5	28.3	
	womleg_2015	state_name					
	<dbl>	<chr>					
1	28.3	alaska					
2	14.3	alabama					
3	20	arkansas					
4	35.6	arizona					
5	25.8	california					

i 45 more rows