

Geography 176A Lab 1

Meet R, RMarkdown, & Github

Aine Roonan

09 August, 2020

Objectives:

- Fork a Git Repo
- Meet RMarkdown
- See some basic data

Your Assignment:

1. Read this document in both knit and raw Rmd format to review Rmarkdown syntax.
2. Change the author in the YML to your name
3. Modify the city threshold to 100,000 (no comma!!)
4. Re-knit the document
5. From the knit document, right click on the final map image and save it to the `img` directory.
6. Push your changes back to your repo
7. Submit the URL to your repo and the saved image to the Guachospace Dropbox.

Data: USA Cites

Accessed 07/27/2020 from: <https://simplemaps.com/data/us-cities>

What is Rmarkdown:

R Markdown provides an unified authoring framework for data science, combining code, results, and commentary. R Markdown documents are fully reproducible and support dozens of output formats, like PDFs, Word files, slideshows, and more.



It contains three important types of content:

1. An (optional) YAML header surrounded by `---` which defines the:
 - Output type
 - theme
 - metadata (author/date/title. subtitle)
2. Text mixed with simple text formatting like `#` heading and *italics*.

3. Chunks of code surrounded by “`”.

How do I make this thing run?

Once, and only when, the document is saved ... Hit,

Cmd/Ctrl + SHIFT + K

or the `knit` button above.

Where do I see the output?

The document will knit to the same directory as the Rmd in the format defined.

Click the settings (wheel) icon next to `knit` to render the output document in a new window (Preview in Window) or in the RStudio Viewer Pane (Preview in Viewer Pane)

RMarkdown formatting basics

Block Quote:

R Markdown provides an authoring framework for data science. You can use a single R Markdown file to both

Header 1 (H1)

Header 2 (H2)

Header 2 (H3)

Header 4 (H4)

Header 5 (H5)

- a bulletpointed
 - list of
 - things
1. Or a
 2. numbered list
 3. of things
 - a. Maybe with some nested sections
 - b. like this
 4. Then something else...

I can make something **bold** with double asterisks, or *italicized* with single asterisks.

Make something superscripted^{UP} or subscripted_{DOWN}

I can add a regular hyperlink by just writing the URL: <https://github.com/mikejohnson51/spds>

Or I can have linked text.

I can add an image:



And I can refine the image:



I can also add document breaks:

Code

These are teasers for where we are going. If the code doesn't yet make sense that is expected!

Focus on what we are doing with Rmarkdown... the exposure to this code and workflow will help when we get to discussing them next week.

All code is run in code snippets. Code snippets can be auto generated with:

Cmd/Ctrl + Alt + I

or by clicking "insert"

The following table summarizes which types of output each option **suppresses**:

Option	Run code	Show code	Output	Plots	Messages	Warnings
eval = FALSE	X		X	X	X	X
include = FALSE		X	X	X	X	X
echo = FALSE		X				
results = "hide"			X			
fig.show = "hide"				X		
message = FALSE					X	
warning = FALSE						X

Inline code:

Read in our data:

```
# load tidyverse
library(tidyverse)

# read in city data
cities = readr::read_csv('../data/uscities.csv')
```

Exploration

Here we see some basic data exploration tools to see what our data looks like:

```
names(cities)
```

```
## [1] "city"          "city_ascii"    "state_id"      "state_name"
## [5] "county_fips"   "county_name"   "county_fips_all" "county_name_all"
## [9] "lat"          "lng"          "population"     "density"
## [13] "source"        "military"      "incorporated"   "timezone"
## [17] "ranking"       "zips"          "id"
```

```
head(cities)
```

```
## # A tibble: 6 x 19
##   city city_ascii state_id state_name county_fips county_name county_fips_all
##   <chr> <chr>      <chr>   <chr>      <dbl> <chr>      <chr>
## 1 Sout~ South Cre~ WA      Washington 53053 Pierce    53053
## 2 Rosl~ Roslyn     WA      Washington 53037 Kittitas 53037
## 3 Spra~ Sprague    WA      Washington 53043 Lincoln  53043
## 4 Gig ~ Gig Harbor WA      Washington 53053 Pierce    53053
## 5 Lake~ Lake Cass~ WA      Washington 53061 Snohomish 53061
## 6 Teni~ Tenino     WA      Washington 53067 Thurston 53067
## # ... with 12 more variables: county_name_all <chr>, lat <dbl>, lng <dbl>,
## #   population <dbl>, density <dbl>, source <chr>, military <lgl>,
## #   incorporated <lgl>, timezone <chr>, ranking <dbl>, zips <chr>, id <dbl>
```

```
summary(cities)
```

```
##      city          city_ascii      state_id      state_name
## Length:28889      Length:28889      Length:28889      Length:28889
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##   county_fips   county_name      county_fips_all   county_name_all
## Min.   : 1001   Length:28889      Length:28889      Length:28889
## 1st Qu.:17187   Class :character  Class :character  Class :character
## Median :30017   Mode  :character  Mode  :character  Mode  :character
## Mean   :29898
## 3rd Qu.:42043
## Max.   :72153
##
##   lat          lng          population      density
## Min.   :17.96   Min.   : -176.63   Min.   :      0   Min.   :    0.0
## 1st Qu.:35.17   1st Qu.: -98.30   1st Qu.:    323   1st Qu.:   115.0
## Median :39.38   Median : -90.20   Median :    1065   Median :   297.0
## Mean   :38.66   Mean    : -92.93   Mean    :   13905   Mean    :   513.7
## 3rd Qu.:41.81   3rd Qu.: -81.77   3rd Qu.:   4304   3rd Qu.:   600.0
```

```
## Max. :70.64 Max. : 173.12 Max. :19354922 Max. :32085.0
## source military incorporated timezone
## Length:28889 Mode :logical Mode :logical Length:28889
## Class :character FALSE:28802 FALSE:8290 Class :character
## Mode :character TRUE :87 TRUE :20599 Mode :character
##
##
## ranking zips id
## Min. :1.000 Length:28889 Min. :1.630e+09
## 1st Qu.:3.000 Class :character 1st Qu.:1.840e+09
## Median :3.000 Mode :character Median :1.840e+09
## Mean :2.947 Mean :1.838e+09
## 3rd Qu.:3.000 3rd Qu.:1.840e+09
## Max. :3.000 Max. :1.840e+09
```

```
str(cities)
```

```
## tibble [28,889 x 19] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ city : chr [1:28889] "South Creek" "Roslyn" "Sprague" "Gig Harbor" ...
## $ city_ascii : chr [1:28889] "South Creek" "Roslyn" "Sprague" "Gig Harbor" ...
## $ state_id : chr [1:28889] "WA" "WA" "WA" "WA" ...
## $ state_name : chr [1:28889] "Washington" "Washington" "Washington" "Washington" ...
## $ county_fips : num [1:28889] 53053 53037 53043 53053 53061 ...
## $ county_name : chr [1:28889] "Pierce" "Kittitas" "Lincoln" "Pierce" ...
## $ county_fips_all: chr [1:28889] "53053" "53037" "53043" "53053" ...
## $ county_name_all: chr [1:28889] "Pierce" "Kittitas" "Lincoln" "Pierce" ...
## $ lat : num [1:28889] 47 47.3 47.3 47.3 48.1 ...
## $ lng : num [1:28889] -122 -121 -118 -123 -122 ...
## $ population : num [1:28889] 2500 947 441 9507 3591 ...
## $ density : num [1:28889] 125 84 163 622 131 491 191 112 50 156 ...
## $ source : chr [1:28889] "polygon" "polygon" "polygon" "polygon" ...
## $ military : logi [1:28889] FALSE FALSE FALSE FALSE FALSE FALSE ...
## $ incorporated : logi [1:28889] TRUE TRUE TRUE TRUE TRUE TRUE ...
## $ timezone : chr [1:28889] "America/Los_Angeles" "America/Los_Angeles" "America/Los_Angeles" ...
## $ ranking : num [1:28889] 3 3 3 3 3 3 3 3 3 ...
## $ zips : chr [1:28889] "98580 98387 98338" "98941 98068 98925" "99032" "98332 98335" ...
## $ id : num [1:28889] 1.84e+09 1.84e+09 1.84e+09 1.84e+09 1.84e+09 ...
## - attr(*, "spec")=
## .. cols(
## .. city = col_character(),
## .. city_ascii = col_character(),
## .. state_id = col_character(),
## .. state_name = col_character(),
## .. county_fips = col_double(),
## .. county_name = col_character(),
## .. county_fips_all = col_character(),
## .. county_name_all = col_character(),
## .. lat = col_double(),
## .. lng = col_double(),
## .. population = col_double(),
## .. density = col_double(),
## .. source = col_character(),
## .. military = col_logical(),
## .. incorporated = col_logical(),
```

```
## .. timezone = col_character(),
## .. ranking = col_double(),
## .. zips = col_character(),
## .. id = col_double()
## .. )
```

```
dim(cities)
```

```
## [1] 28889    19
```

Some basic Data Exploration

Which an understanding of the data, lets answers a few questions:

```
# Which cities are the most dense?
```

```
cities %>%
  select(city, state_id, population, density) %>%
  arrange(-density) %>%
  head()
```

```
## # A tibble: 6 x 4
```

	city	state_id	population	density
	<chr>	<chr>	<dbl>	<dbl>
## 1	Friendship Heights Village	MD	5051	32085
## 2	Manhattan	NY	1643734	27799
## 3	Guttenberg	NJ	11695	23394
## 4	Union City	NJ	70387	21116
## 5	West New York	NJ	54227	21057
## 6	Hoboken	NJ	55131	17026

```
# Which cities have the most people?
```

```
cities %>%
  select(city, state_id, population, density) %>%
  arrange(-population) %>%
  head()
```

```
## # A tibble: 6 x 4
```

	city	state_id	population	density
	<chr>	<chr>	<dbl>	<dbl>
## 1	New York	NY	19354922	11083
## 2	Los Angeles	CA	12815475	3295
## 3	Chicago	IL	8675982	4612
## 4	Miami	FL	6381966	4969
## 5	Dallas	TX	5733259	1524
## 6	Philadelphia	PA	5637884	4547

In line code:

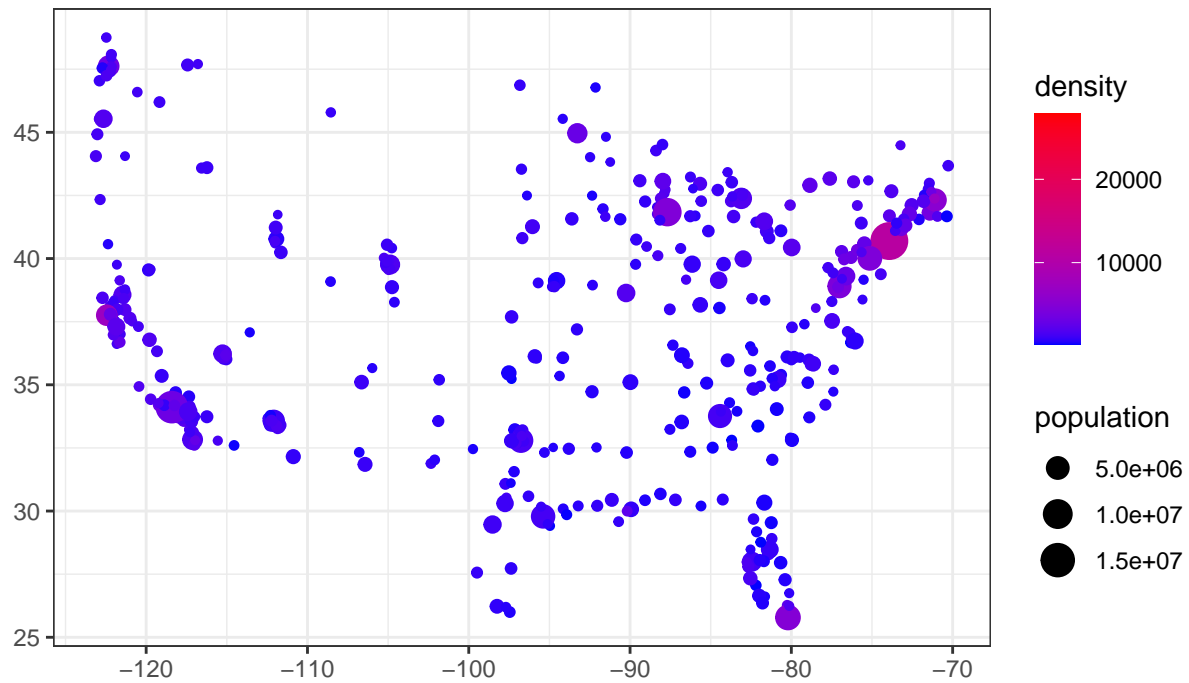
We can evaluate R variables and code in-line:

For example, there are 445 cities with a population greater then 100,000.

Here is a map!

USA Cities Population Stats

Cities bigger then: 100,000 people



Made by: Aine Roonan – Geography 176A