Notes Unit 10
DATA 110
Saidi

Geographic Information Science and Systems

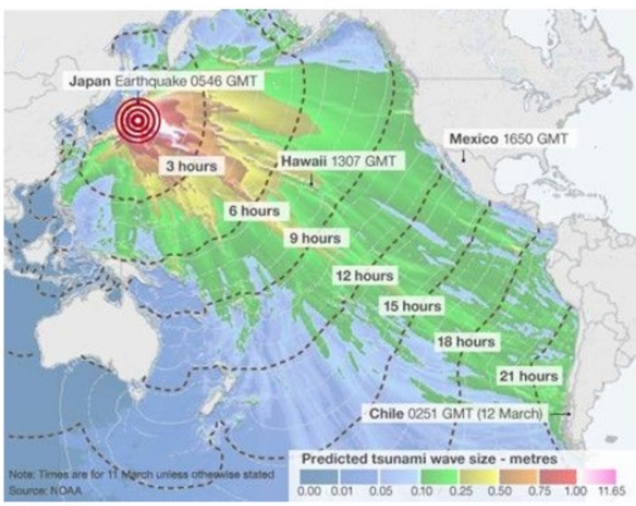
To inspire you, here is what a former student created as her impressive final course project, which incorporates GIS in a heatmap of NYC Asbestos Complaints: https://rpubs.com/rsaidi/617763

What is GIS?

We travel over the Earth, through the lower levels of the atmosphere, and through tunnels below the surface. We dig ditches and bury pipelines and cables, construct mines, drill wells, track pollution, seismic activity, transportation of people and commerce, political activity, social justice issues, environmental issues, and much more. Keeping track of all of this activity and knowing where it occurs can be the most convenient basis for tracking.

The Importance of Location

- Tōhoku earthquake and tsunami
- GI science and systems are integral to response and recovery



Source: NOAA

Geographic Information Science and Systems (Fourth Edition) | Paul A. Longley | Michael F. Goodchild | David J. Maguire | David W. Rhind

The Technology of Problem Solving

GI technology is a combination of:

- a software product, acquired to perform well-defined functions (***GI software***)
- digital representations of aspects of world (***GI***)
- a community of people who use these tools for various purposes (the ***GI community***)
- the **activity** of using GI systems to solve problems or advance science (***'doing GIS'***)

Reminder of Database Volumes (from Week 1 Notes)

Potential GI database volumes in bytes for some typical applications (volumes estimated to the nearest order of magnitude). Strictly, bytes are counted in powers of 2 – 1 kilobyte is 1024 bytes, not 1000.		
1 megabyte	1 000 000 (2^{20})	Single data set in a small project database
1 gigabyte	1 000 000 000 (2^{30})	Entire street network of a large city or small country
1 terabyte	1 000 000 000 000 (2^{40})	Elevation of entire Earth surface recorded at 30 m intervals
1 petabyte	1 000 000 000 000 000 (2^{50})	Satellite image of entire Earth surface at 1 m resolution
1 exabyte	1 000 000 000 000 000 000 (2^{60})	A possible 3-D representation of the entire Earth at 10 m resolution
1 zettabyte	1 000 000 000 000 000 000 000 (2^{70})	One-fifth of the capacity (in 2013) of U.S. National Security Agency Utah Data Center

The Government Perspective

One of the most profound changes concerning government and GI in recent years is in the ways in which government information is shared with citizens and external organizations in many parts of the world.

Public-sector information (PSI) is defined by the wide range of information that government bodies **collect, produce, reproduce, and disseminate** across various areas of activity to fulfill their public-task functions. Many governments worldwide in recent years have committed to greater availability of Open Data – which can be defined as data that can be used, reused, and redistributed freely by anyone.

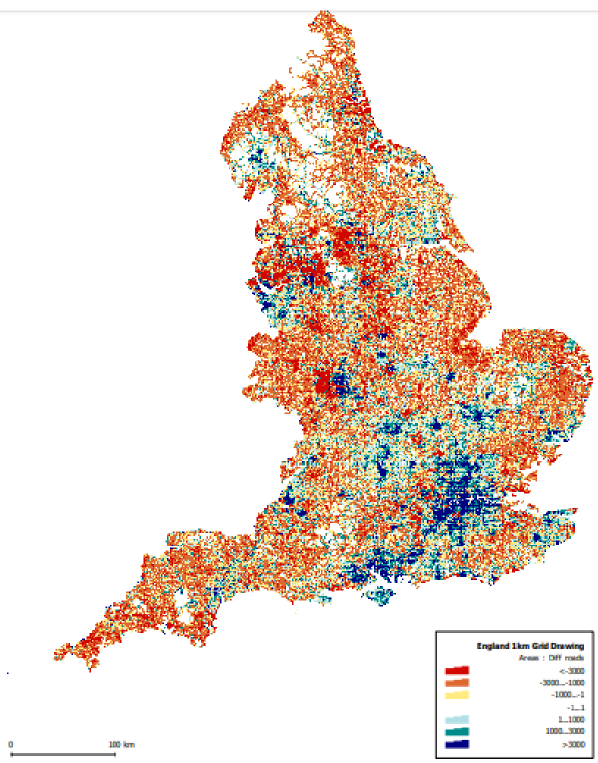
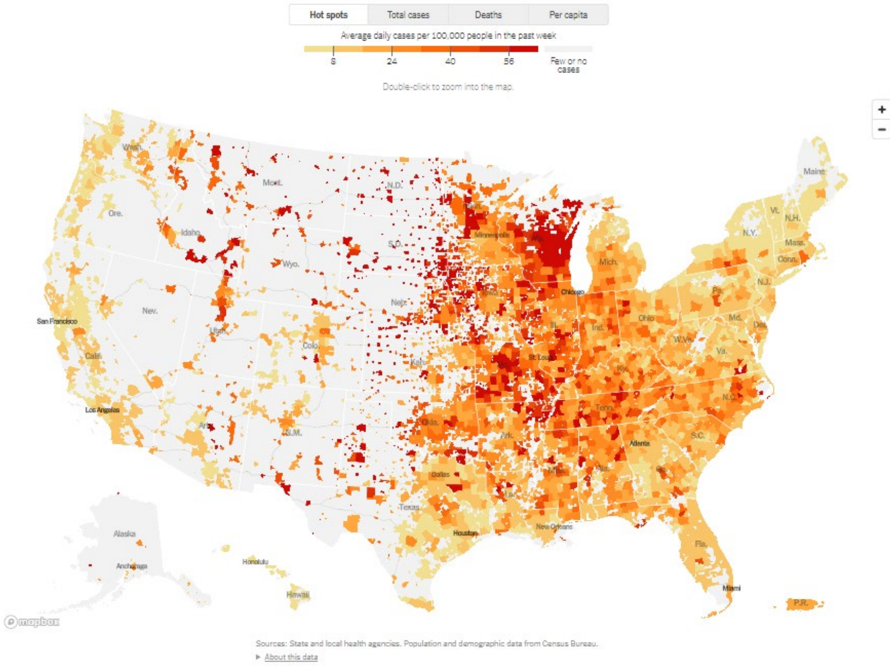


Figure 4 – Length difference between OSM and Meridian datasets. Areas of good OSM coverage are in black, and where it is lacking are in grey.

One major responsibility of governments is disaster management

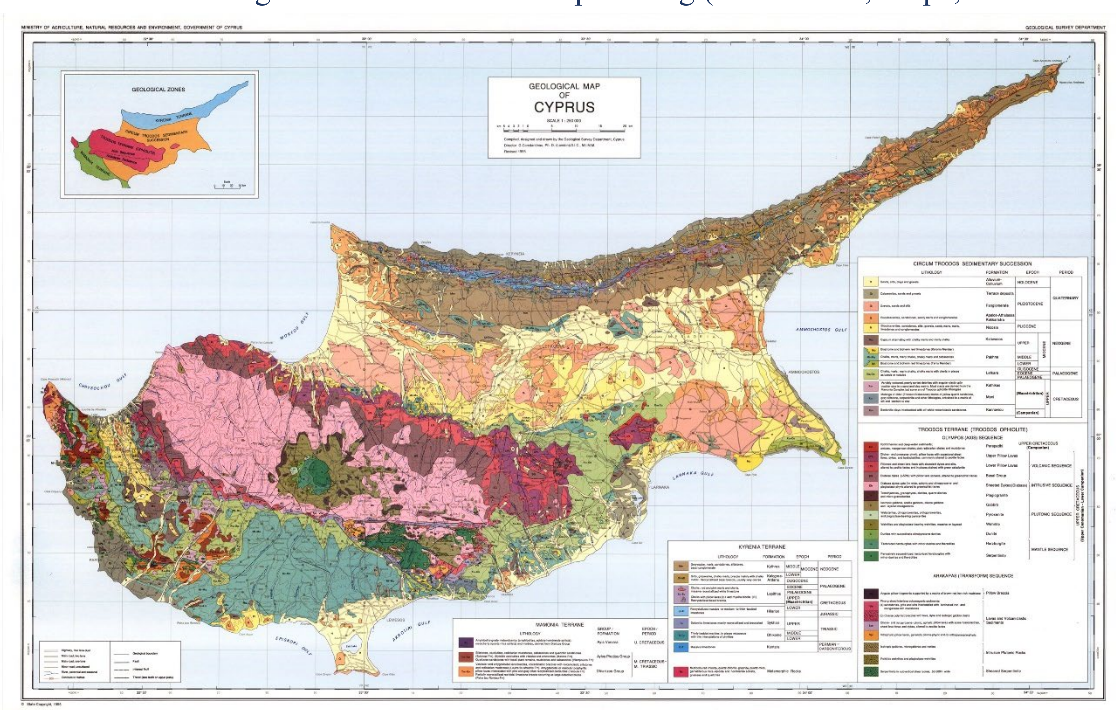
The three Ps of disaster management are: **prevention, preparedness, and protection**. GI Science and Systems are integral to each of them.

We are all familiar with graphs such as this heatmap for hotspots of COVID19 across the US:



https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html

Another obvious government role is in map making (use of color, shape, and size to distinguish characteristics)



Bertin’s Graphic Primatives

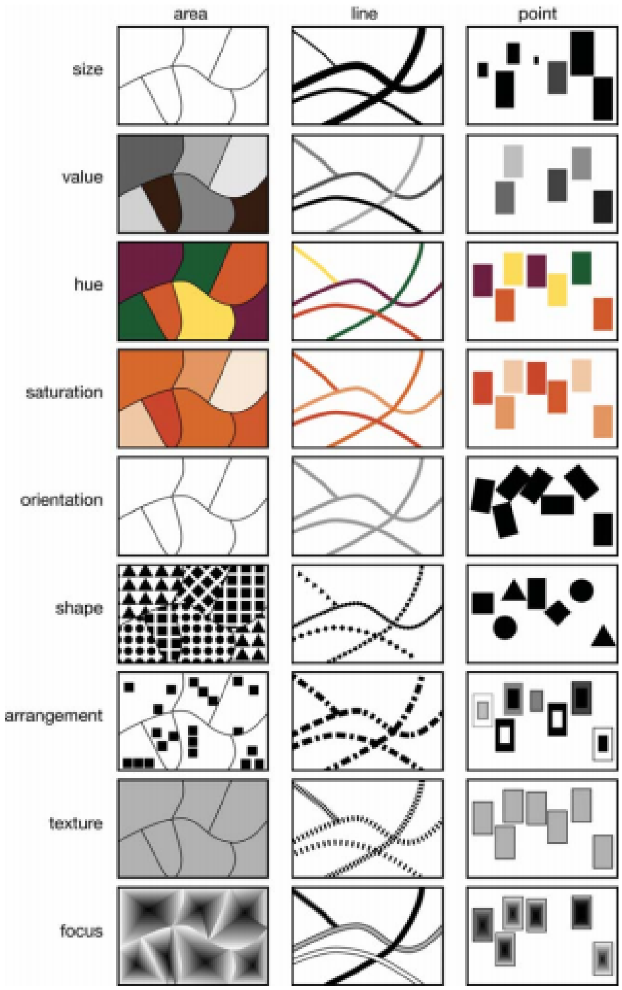
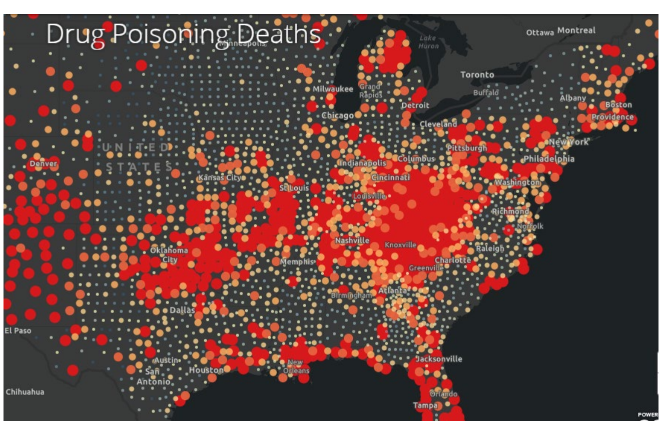


Figure 12.9 Bertin's graphic primitives, extended from seven to ten variables (the variable location is not depicted). *Source:* MacEachren 1994, from *Visualization in Geographical Information Systems*, Hearnshaw H.M. and Unwin D.J. (eds), Plate B. (Reproduced by permission of John Wiley & Sons, Ltd.)

Geovisualization (Uses, Users, Messages, and Media)

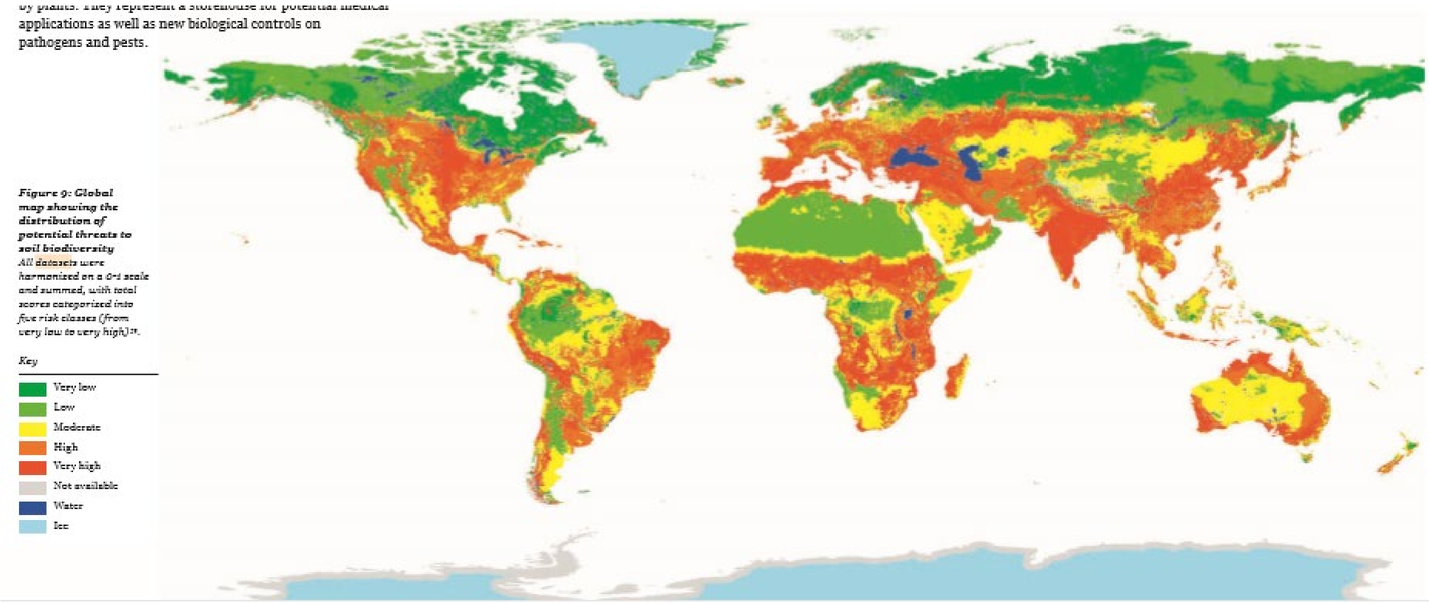
Here is a map-related visualization I came across on the US opioid epidemic, by ESRI. It utilizes ArcGIS software. Be patient – there is quite a bit of interactivity with this graph.

[The Opioid Epidemic]

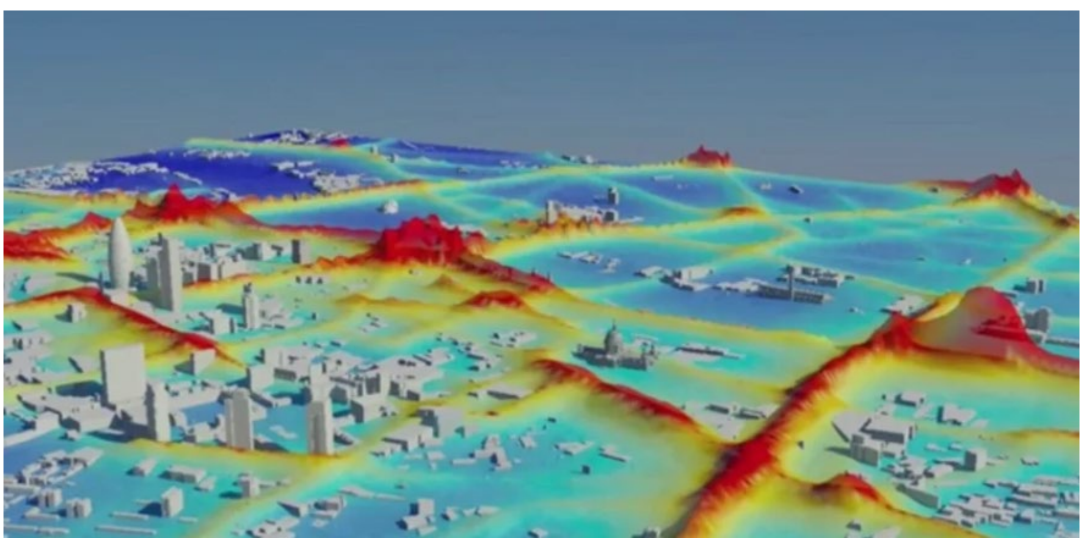


Below are two static, more traditional visualizations from the Lonely Planet Report 2018 as a snapshot of consumption worldwide.

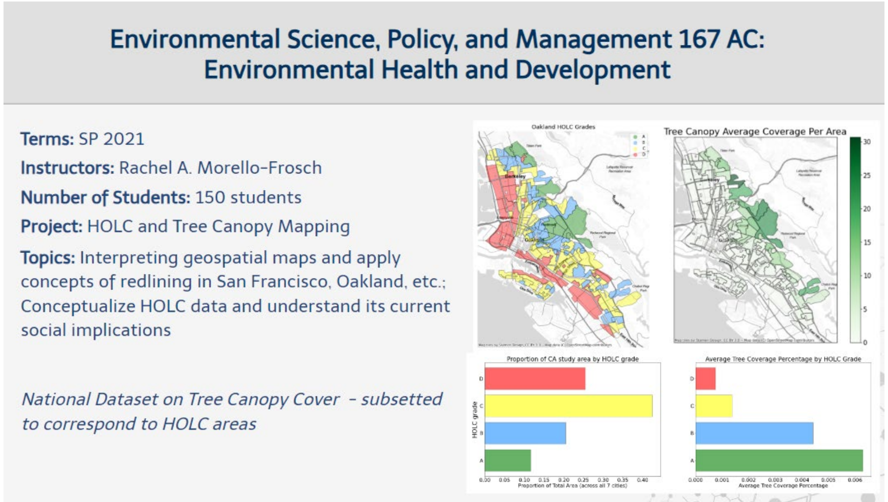
This geovisualization by the Lonely Planet report describes potential threats to soil biodiversity.



The graph below shows a visualization of nitrogen oxide (NOx) pollution in the Virtual London model. This pollutant, often called "urban smog" is largely derived from vehicle emissions: red identifies higher levels, whereas blue represents lower levels. Common analysis for measuring these pollutants are infrared, chemiluminescence and electrochemical.



Using GIS Data to Address Disparities



Creating Maps in R

Many types of software exist to help map data easily, such as ArcGIS from ESRI, Tableau (and Tableau Public), and QGIS. It is not an easy task to make maps in R and takes many steps. The most difficult tasks are: ensuring the data is in the correct format, finding the correct "shape" data (geo information for your data set), merging the two data sets. After these tasks are complete, mapping is not challenging.

Shapefiles can be obtained from the US Census Bureau under the [TIGER products](). There are many options for types of shapefiles to choose from. In your tutorial, you will be directed as to which file to download.

In case you want to go to the correct file directly in TIGER, use this link: https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.2014.html

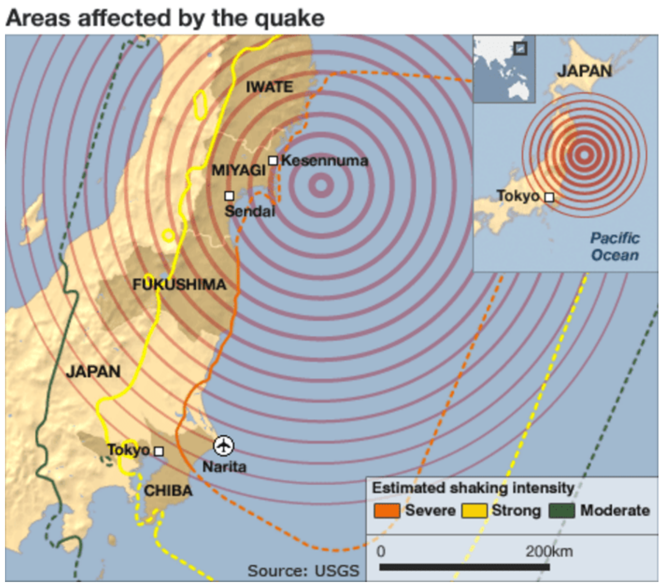
One other concept you will need is the FIPS Code, or the **Federal Information Processing Standard publication**. Counties and states all have FIPS codes.

State FIPS codes:
https://www.census.gov/library/reference/code-lists/ansi.html#state

County FIPS codes:
https://en.wikipedia.org/wiki/List_of_United_States_FIPS_codes_by_county

Japan Quakes

Japan Earthquakes



2011

Geo mapping using Japan Earthquakes 2001-2018 dataset. Source: USGS Earthquake Catalog (https://earthquake.usgs.gov/earthquakes/search/).

This tutorial is adapted from: https://towardsdatascience.com/how-to-make-stunning-geomaps-in-r-a-complete-guide-with-leaflet-be1b857f1644

This is a large dataset with over 14,000 observations and 22 variables, including:

Variables Used
time
longitude
latitude
depth
mag (magnitude)
magType (magnitude type)
place
type
horizontalError
depthError
magError

Load the necessary libraries

```
library(leaflet)
library(sf)
```

```r
library(tidyverse)
library(knitr)
```

Set the lat and long values for Japan

```r
japan_lat <- 138.129731
japan_lon <- 38.0615855
```

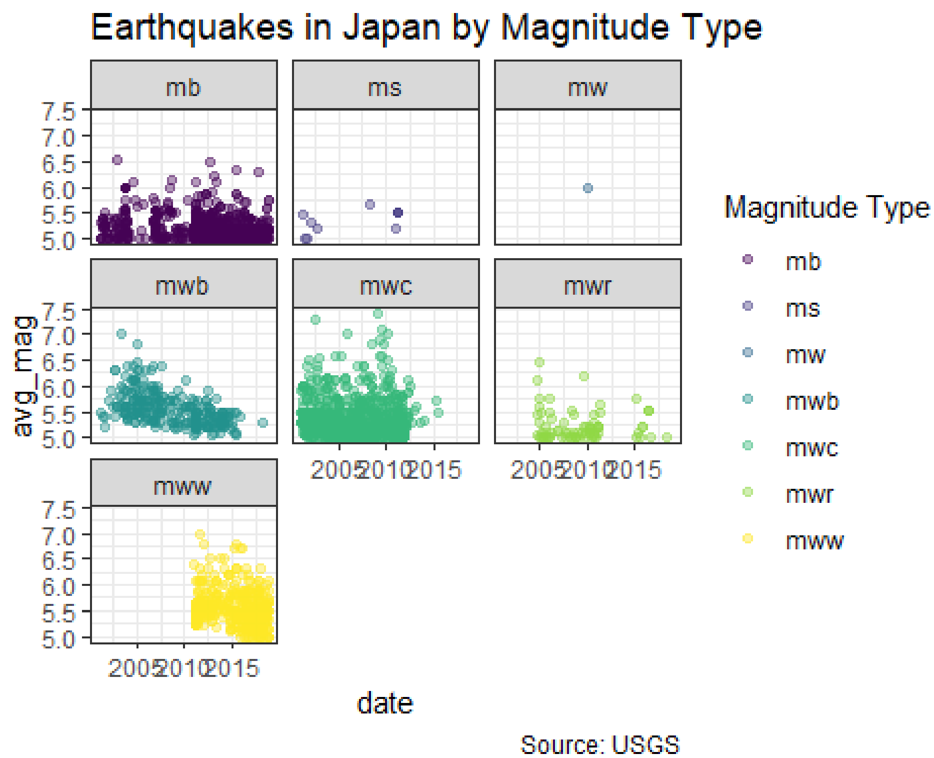
Set the working directory and read in the earthquake data

```r
setwd("C:/Users/rsaidi/Dropbox/Rachel/MontColl/Datasets/Datasets")
quakes <- read_csv("japan_quakes_01-18.csv")
```

Look at Japan’s earthquakes over time

First create a new variable date that removes the time stamp from the time variable

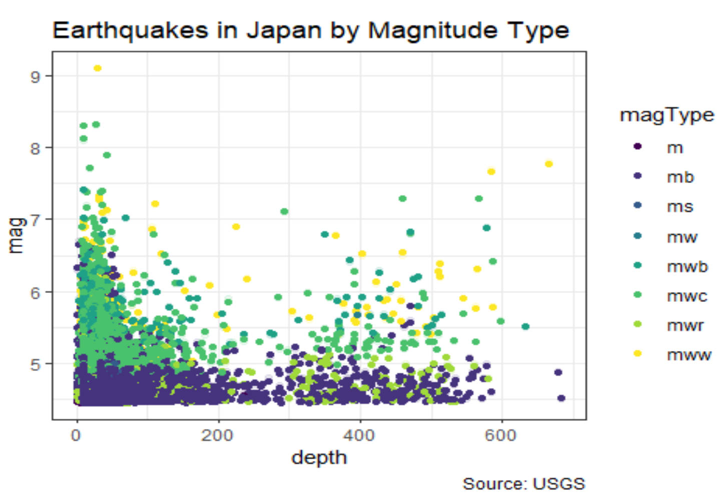
```r
library(lubridate)
quakes$date <- as.Date(format(ymd_hms(quakes$time),format = '%Y-%m-%d'))
head(quakes)

# A tibble: 6 × 23
  time                latitude longitude depth   mag magType   nst   gap  dmin
  <dttm>                 <dbl>     <dbl> <dbl> <dbl> <chr>   <dbl> <dbl> <dbl>
1 2018-11-27 14:34:20     48.4      155.  35     4.9 mb         NA    92 5.04 
2 2018-11-26 23:33:50     36.1      140.  48.8   4.8 mww        NA   113 1.36 
3 2018-11-26 13:04:02     38.9      142.  50.6   4.5 mb         NA   145 1.29 
4 2018-11-26 05:20:16     50.1      156.  66.3   4.6 mb         NA   128 3.19 
5 2018-11-25 09:19:05     34.0      134.  38.2   4.6 mb         NA   104 0.558
6 2018-11-25 03:16:46     48.4      155.  35     4.6 mb         NA    85 4.99 
#
```



Here, we can see that mww and mb types were more prevalent after 2011 and prior to 2011, mwc and mwb were more prevalent types.

Create a scatterplot of magnitude versus depth

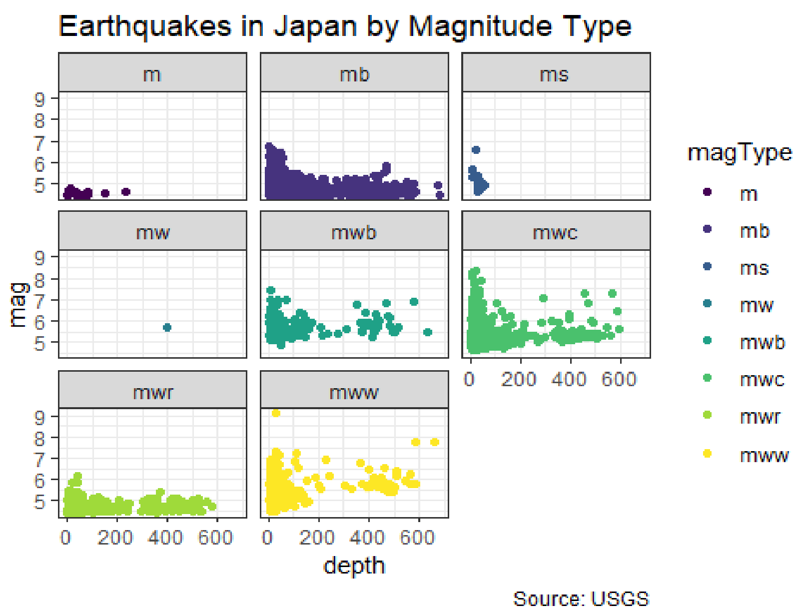
```
ggplot(quakes, aes(x=depth, y=mag, color = magType)) +
  geom_point(alpha = 0.1) +
  scale_color_viridis_d()+
  geom_jitter() +
  labs(title = "Earthquakes in Japan by Magnitude Type",
       caption = "Source: USGS")  +
  theme_bw()
```



This plot appears to have two clumps with peaks at <50 meters depth and about 400-500 meters depth

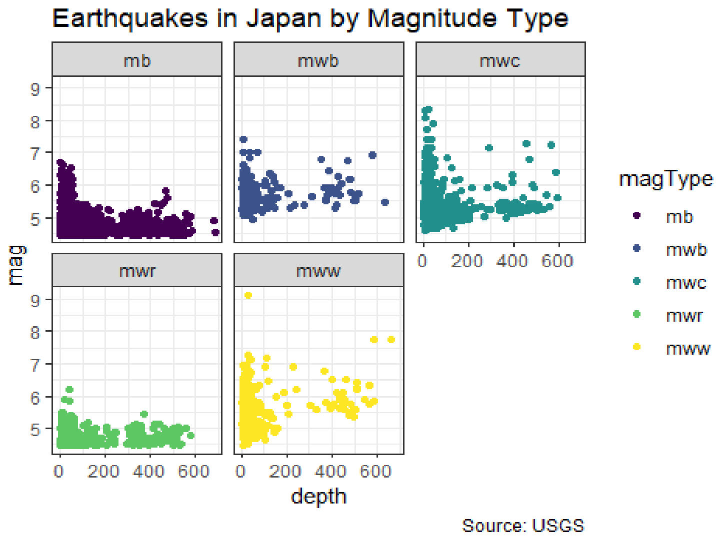
Try to Facet for the magType

```
ggplot(quakes, aes(x=depth, y=mag, color = magType)) +
  geom_point(alpha = 0.05) +
  scale_color_viridis_d()+
  geom_jitter() +
  facet_wrap(~magType) +
  labs(title = "Earthquakes in Japan by Magnitude Type",
       caption = "Source: USGS")  +
  theme_bw()
```



Since there are so few, remove magTypes: m, ms, and mw

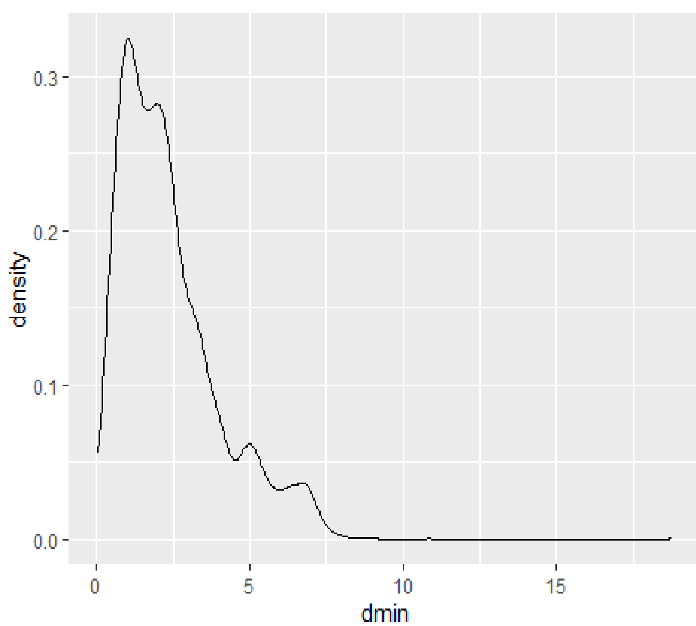
```
quakes2 <- quakes |> filter(magType %in% c("mb", "mwb", "mwc", "mwr", "mww"))
ggplot(quakes2, aes(x=depth, y=mag, color = magType)) +
  geom_point(alpha = 0.05) +
  scale_color_viridis_d()+
  geom_jitter() +
  facet_wrap(~magType) +
  labs(title = "Earthquakes in Japan by Magnitude Type",
       caption = "Source: USGS")  +
  theme_bw()
```



Mwc, mwb, and mww in particular seem to increase in magnitude for depths greater than 400 meters.

Explore the duration

```
ggplot(quakes, aes(dmin)) +
  geom_density(bins = 15)

Warning in geom_density(bins = 15): Ignoring unknown parameters: `bins`

Warning: Removed 10485 rows containing non-finite values (`stat_density()`).
```

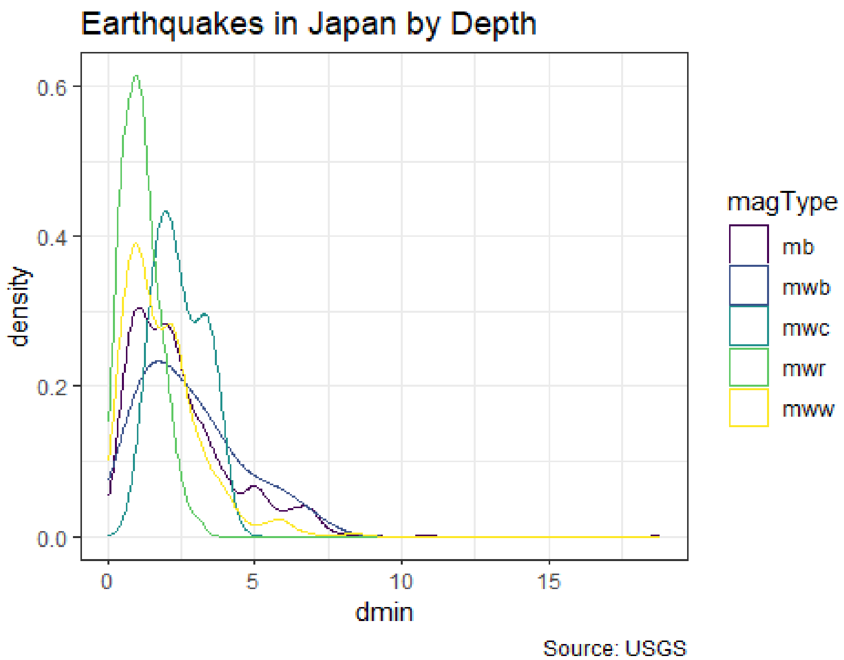


Most earthquakes last for under 7 minutes

Explore the outliers

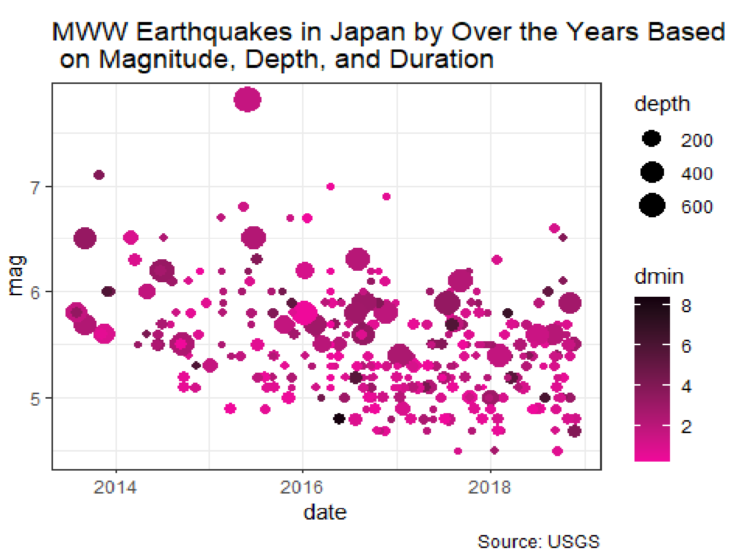
```
ggplot(quakes2, aes(x=dmin, color = magType))+
  geom_density(alpha = 0.4) +
  scale_color_viridis_d()  +
  theme_bw() +
```

```
  labs(title = "Earthquakes in Japan by Depth",
       caption = "Source: USGS")

Warning: Removed 10444 rows containing non-finite values (`stat_density()`).
```



mww has some of the longest durations with the greatest magnitudes and depths

```
quakes_mww <- quakes |>
  filter(magType == "mww") |> filter(!is.na(mag) & !is.na(depth) & !is.na(dmin))
ggplot(quakes_mww, aes(x=date, y=mag, color = dmin)) +
  geom_point(aes(size=depth)) +
  scale_color_gradient(high = "#14010d", low = "#f2079c")  +
  theme_bw() +
  labs(title = "MWW Earthquakes in Japan by Over the Years Based \n on Magnitude, Depth, 
and Duration",
       caption = "Source: USGS")
```



Filter for magnitudes greater than 6

```
strong <- quakes |>
  filter(mag >= 6)
```

Mapping Earthquakes

Decide what style you would like your map

Use this link to look at all the options: https://leaflet-extras.github.io/leaflet-providers/preview/

Draw a first map using leaflet and the Esri World Street Map

Use the function addProviderTiles() to decide which style you want - I have included a few different types for you to try out.

Some examples:

- addProviderTiles("Stamen.Terrain")
- addProviderTiles("Stamen.Watercolor")
- addProviderTiles("Esri.WorldPhysical")
- addProviderTiles("Esri.NatGeoWorldMap")

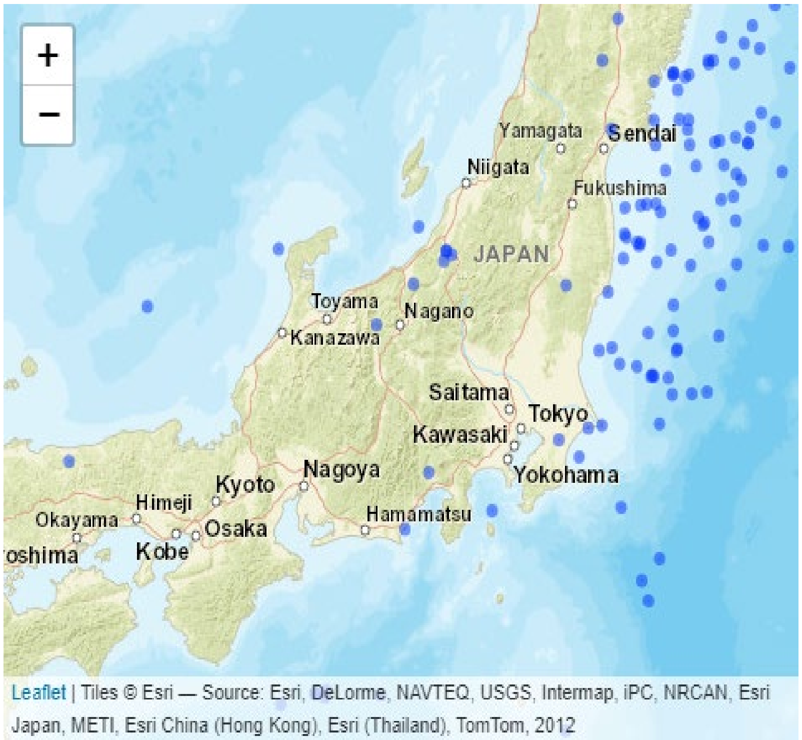
First, we need to set the location of the map of Japan. A quick google search gives 36.2 degrees North, 138.2 degrees East. That translates to lng = 138.2, lat = 36.2

(North = + lat, South = - lat, East = + lng, West = - lng)

Zoom provides the level of granularity. Play with zoom = 1 versus zoom = 6 to see what the initial setting is.

```
leaflet() |>
  setView(lng = 138.2, lat = 36.2, zoom =6) |>
  addProviderTiles("Esri.WorldStreetMap") |>
  addCircles(
    data = strong,
    radius = strong$mag
)
```

```
Assuming "longitude" and "latitude" are longitude and latitude, respectively

PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is insta
lled, please make sure the phantomjs executable can be found via the PATH variable.
```



Tweak the marker size

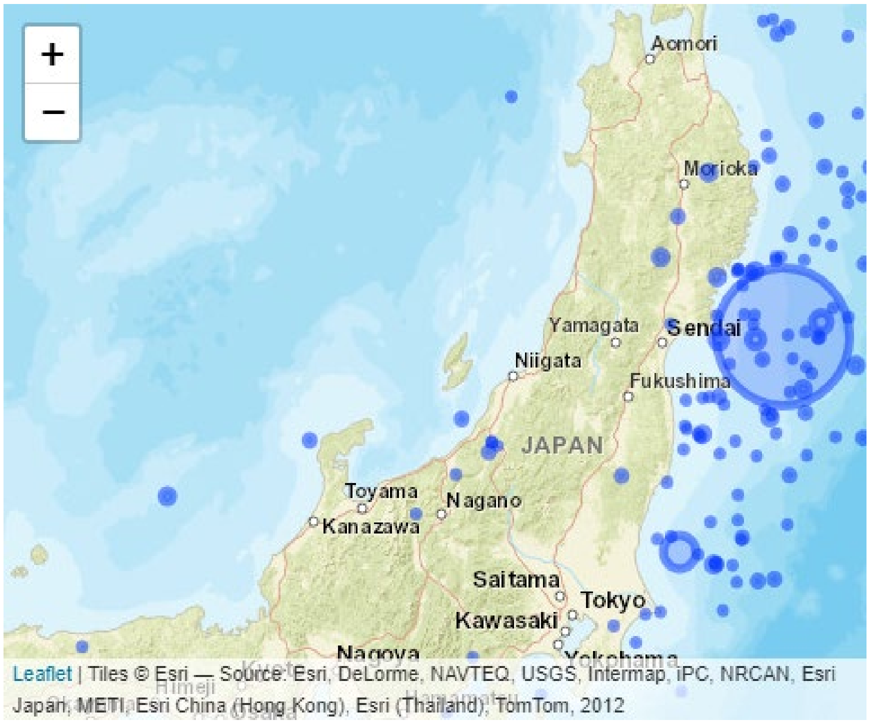
You can use the following completely non-scientific formula to calculate marker size:

$$\sqrt{10^x} \times c$$

The x here represents the magnitude, and c represents a constant you can play around with. The larger it is, the easier it is to spot stronger earthquakes.

Implement this formula in the *radius* with the value for c is set to 2.

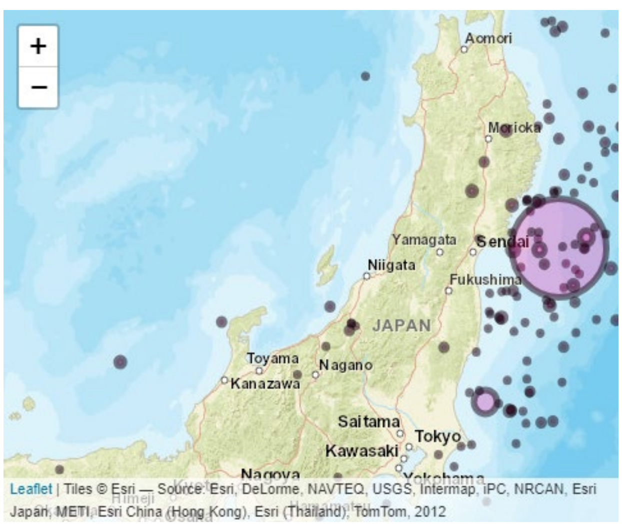
```
leaflet() |>
  setView(lng = japan_lat, lat = japan_lon, zoom = 6) |>
  addProviderTiles("Esri.WorldStreetMap") |>
  addCircles(
    data = strong,
    radius = sqrt(10^strong$mag)*2
  )

Assuming "longitude" and "latitude" are longitude and latitude, respectively
```



Tweak the markers

You can tweak the markers using addCircles() in terms of color, fillColor, and fillOpacity

```
leaflet() |>
  setView(lng = japan_lat, lat = japan_lon, zoom = 6) |>
  addProviderTiles("Esri.WorldStreetMap") |>
  addCircles(
    data = strong,
    radius = sqrt(10^strong$mag) * 2,
    color = "#14010d",
    fillColor = "#f2079c",
    fillOpacity = 0.25
  )

Assuming "longitude" and "latitude" are longitude and latitude, respectively
```



Create a popup using paste0

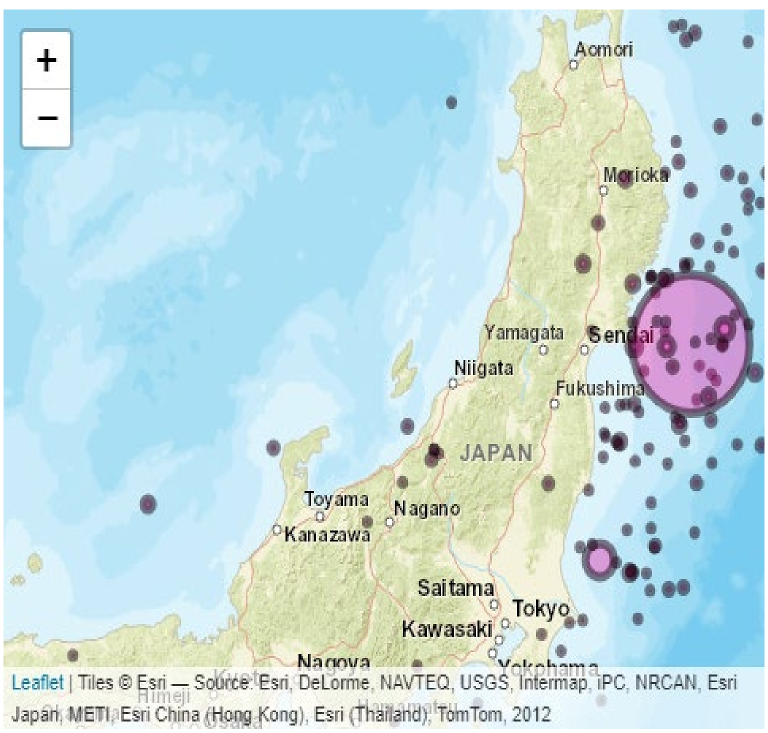
- create a line break using < br >
- surround text with < b > makes it bold

```
popupquake <- paste0(
      "<b>Time: </b>", strong$time, "<br>",
      "<b>Magnitude: </b>", strong$mag, "<br>",
      "<b>Depth (km): </b>", strong$depth, "<br>",
      "<strong>Place: </strong>", strong$place, "<br>"
    )
```

Add the popup to the map

Click on the points to see the popup tooltip for details about each point

```
leaflet() |>
  setView(lng = japan_lat, lat = japan_lon, zoom = 6) |>
  addProviderTiles("Esri.WorldStreetMap") |>
  addCircles(
    data = strong,
    radius = sqrt(10^strong$mag) * 2,
    color = "#14010d",
    fillColor = "#f2079c",
    fillOpacity = 0.35,
    popup = popupquake
  )
```



Notice the largest point is in 2011, with a magnitude of 9.1, which caused the Fukushima nuclear power plant meltdown.

Week 10 Homework

1. **(ungraded)** Complete the Japan Earthquakes Tutorial.

2. **(Worth up to 15 points)** Load the "Healthy Cities GIS Assignment" QMD.
   Complete steps 1 – 5 at the end of the document. Knit or render your document to rpubs and submit in the assignment dropbox by **11:59 pm Tuesday, _____.**

   You will present on Wednesday.