Icahn
School of
Medicine at
Mount
Sinai

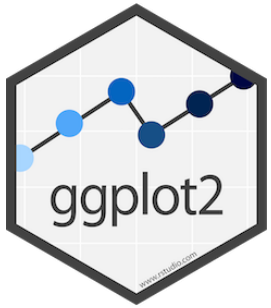
Plotting with ggplot2

Part I

Joselyn Chavez

05/17/2022

The ggplot2 package



- Part of the tidyverse
- Based on "The grammar of graphics" (Leland Wilkinson, 2000)
- Structured syntaxis based on layers
- 110 registered extensions https://exts.ggplot2.tidyverse.org/gallery/
  - gganimate
  - ggthemes
  - ggpubr

Some example plots made with ggplot2

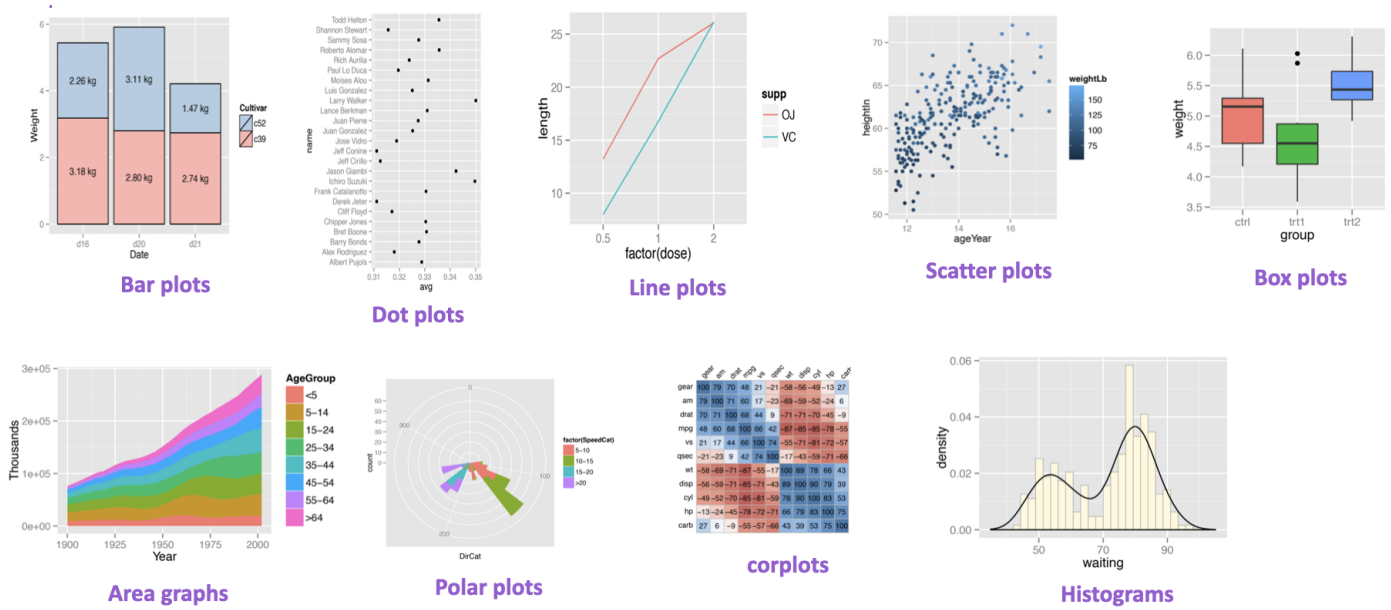
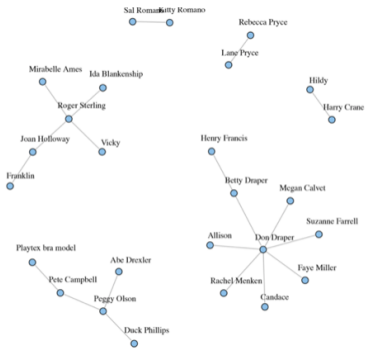
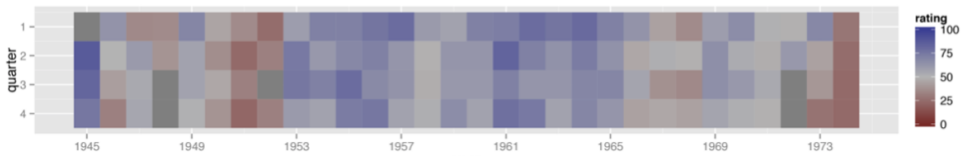


Image from [Godoy, 2021]

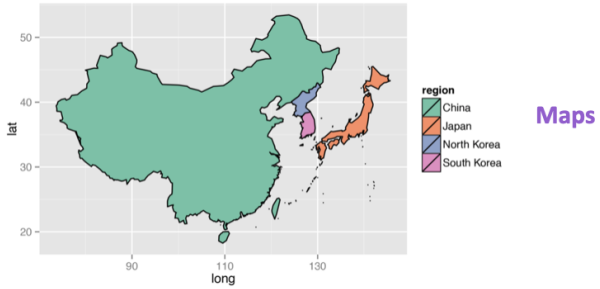
Some example plots made with ggplot2



Network plots



Heatmaps



Maps

Image from [Godoy, 2021]

Major components of the Grammar of Graphics

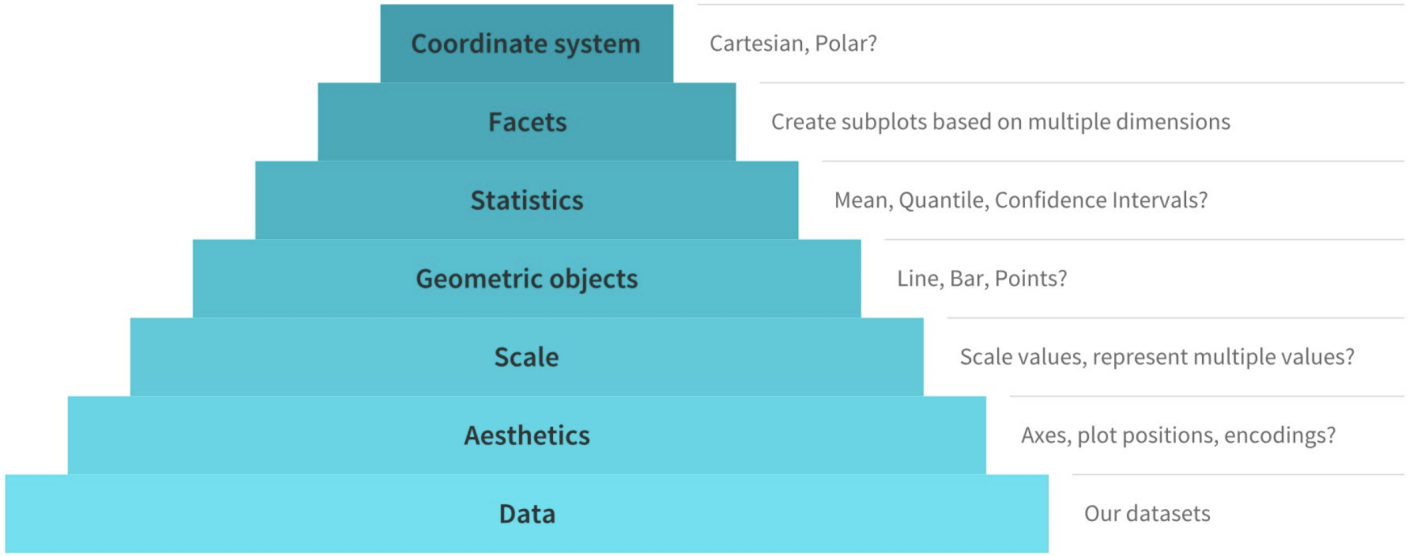


Image from [Carpentry Lessons]

Map and plot data

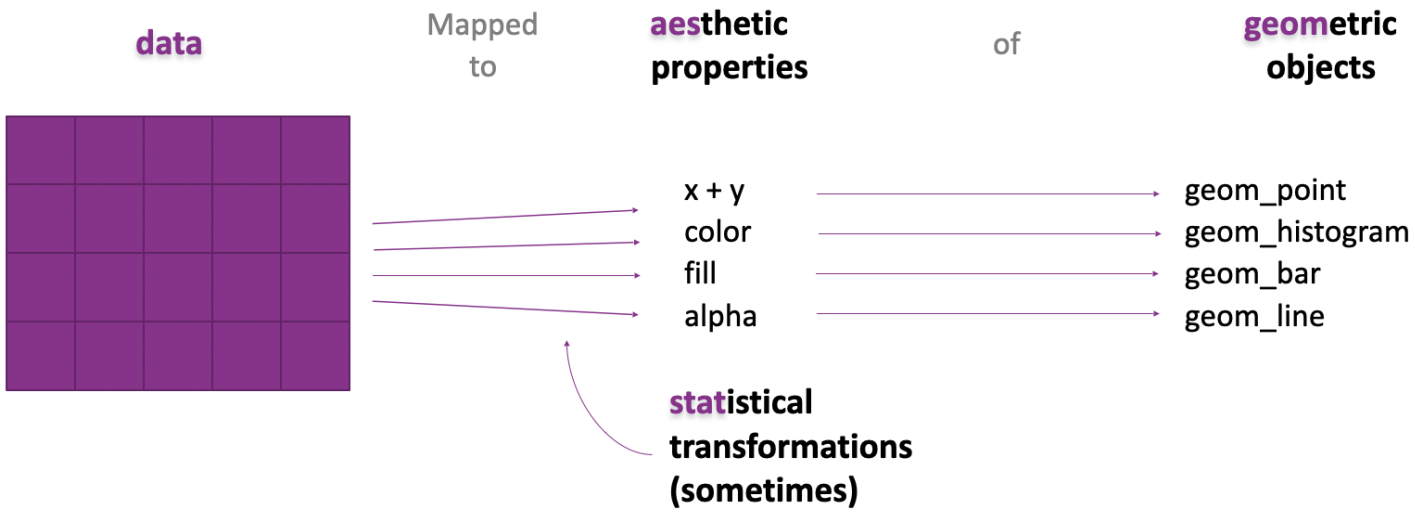
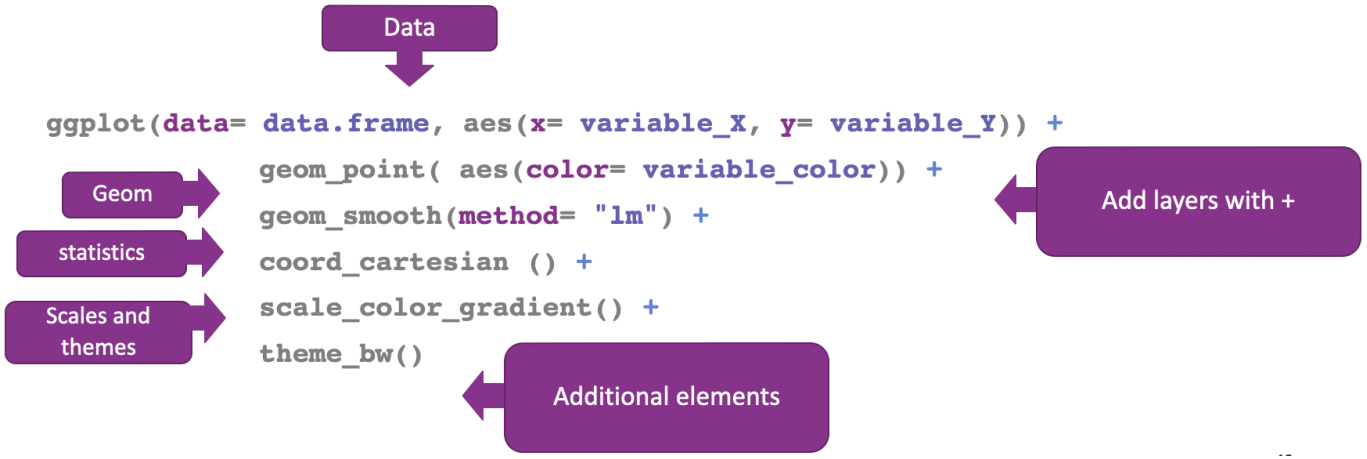


Image from [rfortherestofus]

Syntaxis of ggplot2



Our first ggplot

- Read the data

```
library(tidyverse)
library(janitor)

sinai_covid <- read_csv("sinai_covid.csv")

sinai_covid <- sinai_covid %>%
  clean_names()
```

- ggplot() will start a blank canvas

```
ggplot()
```

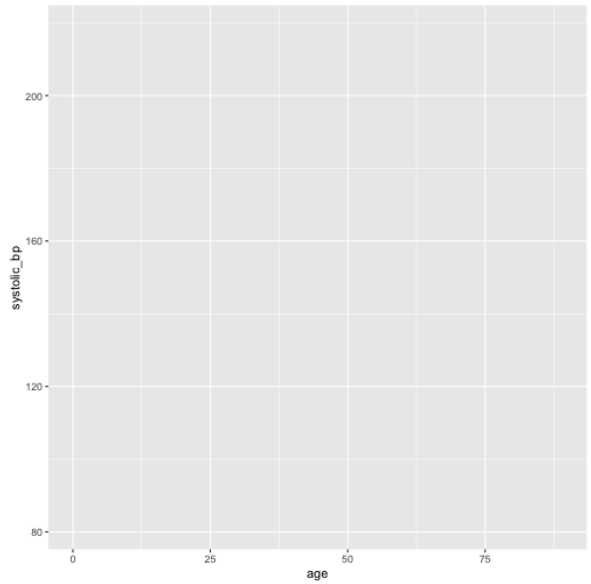
- Add the data to plot

```
ggplot(sinai_covid)
```


Still a blank canvas

- Add the aesthetics properties

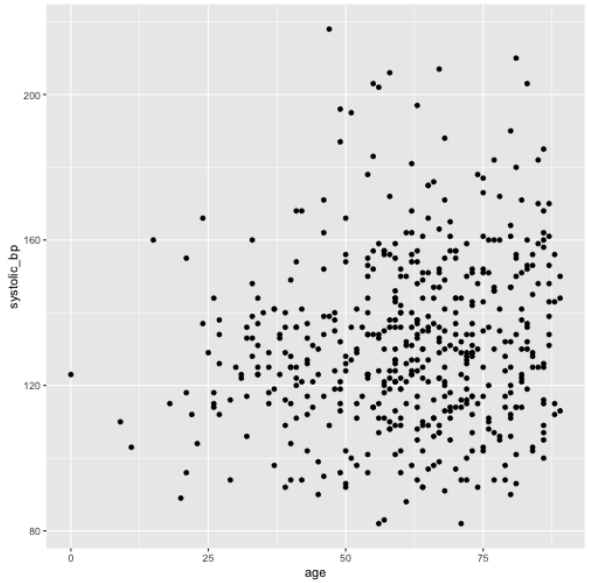
```
ggplot(sinai_covid,
       aes(x = age, y = systolic_bp))
```



Still a blank canvas, with axis

- Add the geometric objects

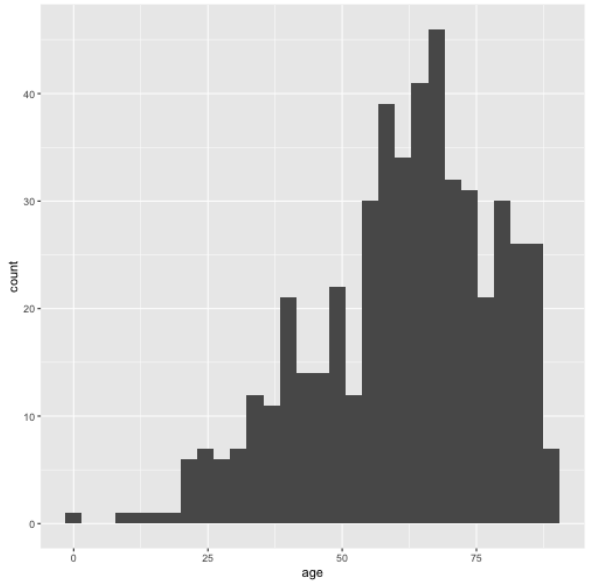
```
ggplot(sinai_covid,
       aes(x = age, y = systolic_bp)) +
  geom_point()
```



We got a scatterplot!

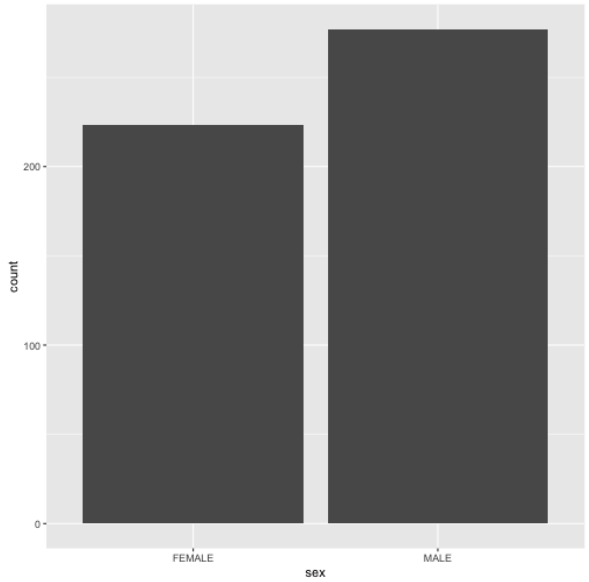
Plotting 1 variable (num)

```
ggplot(sinai_covid, aes(x = age)) +
  geom_histogram()
```



Plotting 1 variable (cat)

```
ggplot(sinai_covid, aes(x = sex)) +
  geom_bar()
```

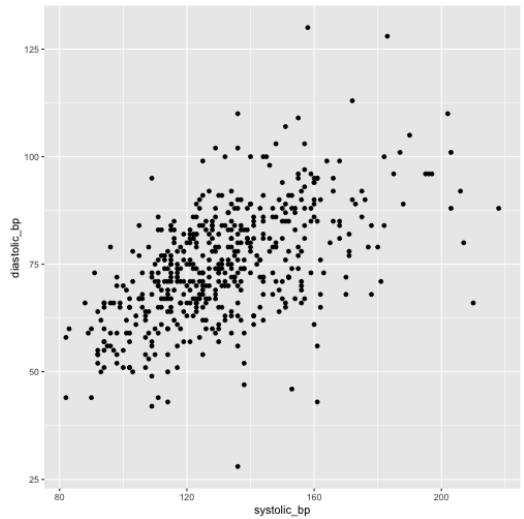


Your turn!

- Plot the distribution of systolic_bp
- Plot the number of patients depending on their smoking status

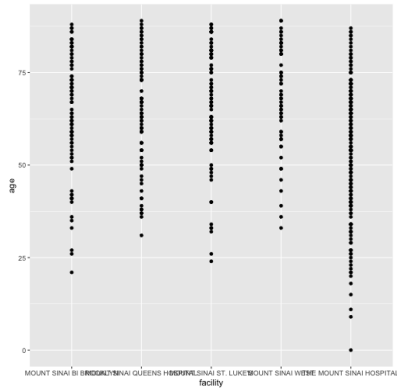
Plotting 2 variables (num-num)

```
ggplot(sinai_covid,
       aes(x = systolic_bp, y = diastolic_bp)
  geom_point()
```



Plotting 2 variables (cat-num)

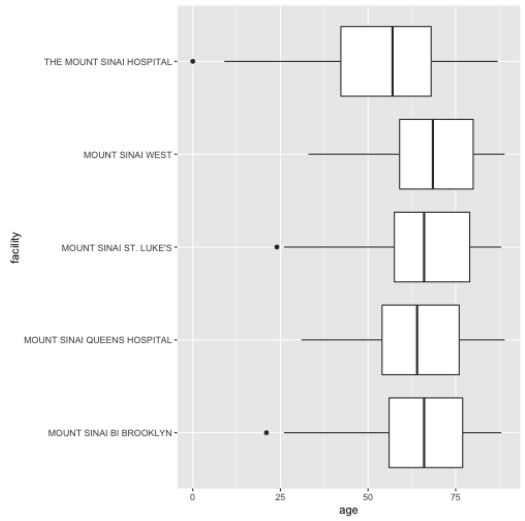
```
ggplot(sinai_covid,
       aes(x = facility, y = age)) +
  geom_point()
```



No so good...

Plotting 2 variables (cat-num) - Much better!

```
ggplot(sinai_covid,
       aes( x = age, y = facility,)) +
  geom_boxplot()
```

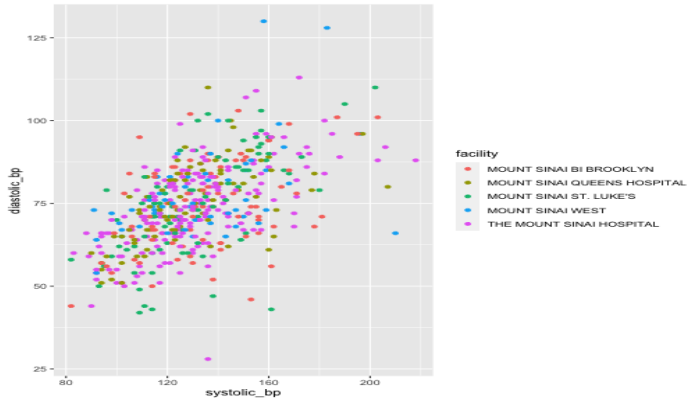


Your turn!

- Plot the variables age vs bmi
- Plot the distribution of patient's bmi from the different facilities

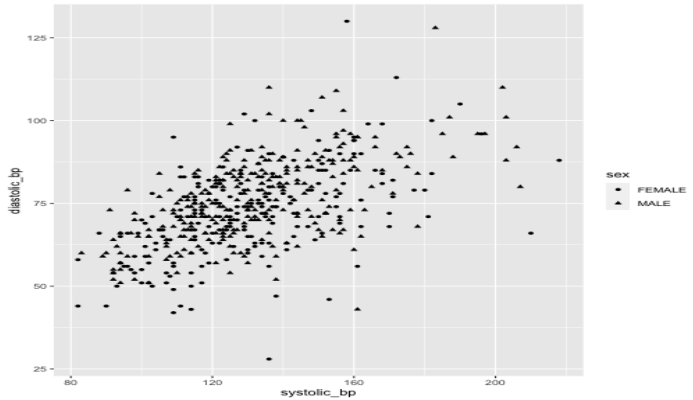
Plotting 3 variables (num-num-cat)

```
ggplot(sinai_covid,
       aes(x = systolic_bp,
           y = diastolic_bp,
           color = facility)) +
  geom_point()
```



Plotting 3 variables (num-num-cat)

```
ggplot(sinai_covid,
       aes(x = systolic_bp,
           y = diastolic_bp,
           shape = sex)) +
  geom_point()
```

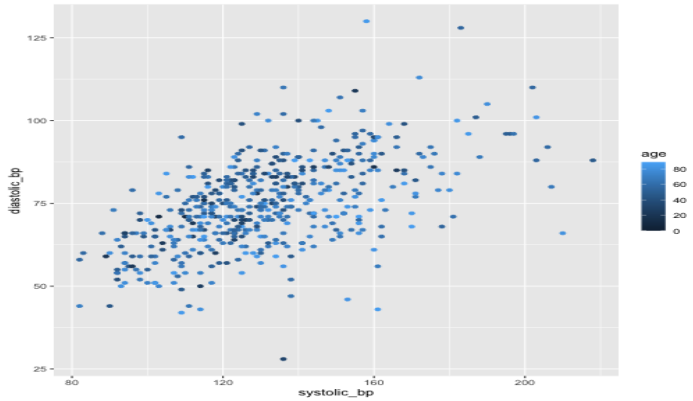


Let's practice

- Plot the patient's age vs bmi, and separate by color or shape based on their smoking status

Plotting 3 variables (num-num-num)

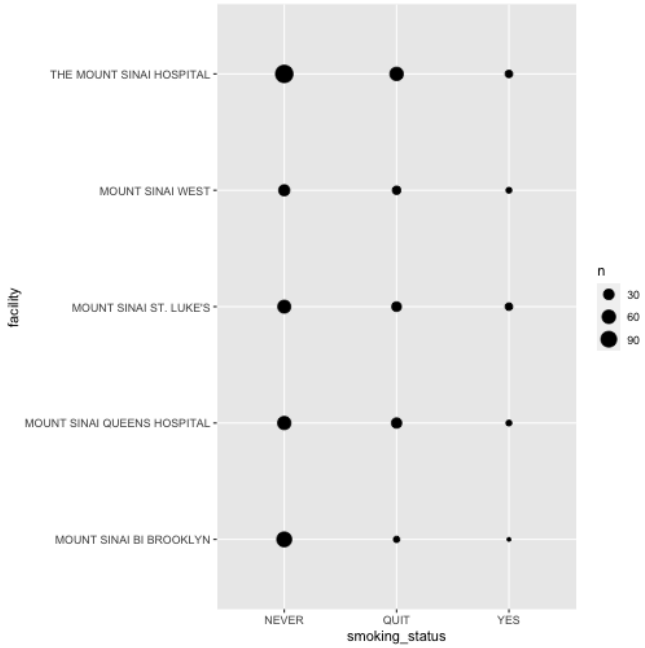
```
ggplot(sinai_covid,
       aes(x = systolic_bp,
           y = diastolic_bp,
           color = age)) +
  geom_point()
```



Plotting 3 variables (cat-cat-num)

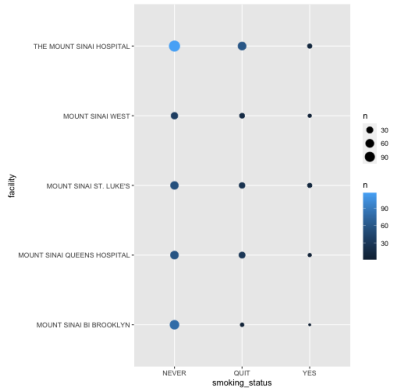
```
count_smoking <- sinai_covid %>%
  group_by(facility) %>%
  count(smoking_status)

ggplot(count_smoking,
       aes(x = smoking_status,
           y = facility,
           size = n)) +
  geom_point()
```

Plotting 3 variables (cat-cat-num)



Plotting 3 variables (cat-cat-num)

```
ggplot(count_smoking,
       aes(x = smoking_status,
           y = facility,
           size = n, color = n)) +
  geom_point()
```



Let's practice

- Count the number of patients with chronic kidney disease per facility and plot the data

Thanks!

Ilustration by Allison Horst