04_BIOS512_assignment

October 31, 2022

1 BIOS512 Assignment #4

Import the Tidyverse into your R session

[9]:
```
library('tidyverse')
library('scales')
```

```
Attaching package: ‘scales’


The following object is masked from ‘package:purrr’:

    discard


The following object is masked from ‘package:readr’:

    col_factor
```

In this assignment we'll be using data from the TidyTuesday Project. Specifically, we'll investigate vaccination rates at US schools. A CSV file of the data is located at:

https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-02-25/measles.csv

Get the "raw" data URL and import it into your R session using `read_csv`. Remember to capture the data under a variable name of your choosing!

[10]:
```
us_vacc = read_csv('https://raw.githubusercontent.com/rfordatascience/
  ↪tidytuesday/master/data/2020/2020-02-25/measles.csv', col_types = cols())
us_vacc |> head()
```

A tibble: 6 × 16

index <dbl>	state <chr>	year <chr>	name <chr>	type <chr>	city <chr>
1	Arizona	2018-19	A J Mitchell Elementary	Public	Nogales
2	Arizona	2018-19	Academy Del Sol	Charter	Tucson
3	Arizona	2018-19	Academy Del Sol - Hope	Charter	Tucson
4	Arizona	2018-19	Academy Of Mathematics And Science South	Charter	Phoenix
5	Arizona	2018-19	Acclaim Academy	Charter	Phoenix
6	Arizona	2018-19	Alfred F Garcia	Public	Phoenix

The `mmr` column has holds measles, mumps, rubella vaccination rates for students in each respective school. **If the `mmr` value is not available for a school the `mmr` value is set to -1 in this data set.**

The target `mmr` vaccination rate as recommended by the CDC for schools is 95%.

Calculate the fraction of schools per state that have vaccination rates greater than the CDC 95%. Capture the output as a table called `df_vacc_rates`.

You'll need to do use `filter`, `group_by`, and `summarize`.

I.e. 1. Filter out schools that don't report `mmr` vaccination rate (keep schools where `mmr >= 0`). 1. Group the data by `state`. 1. Summarize fraction of vaccination rates above 95%.

Remember `n()` stands for the number or records in a group. Also, `sum(mmr >= 95, na.rm=TRUE)` will count the number of values above or equal to 95.

[11]:
```
# create the df_vacc_rates here
# create the df_vacc_rates here
df_vacc_rates = us_vacc |>
    filter(mmr >= 0) |>
    group_by(state) |>
    summarize(frac = sum(mmr >= 95, na.rm = TRUE) / n())
df_vacc_rates
```

state	frac
<chr>	<dbl>
Arizona	0.506404782
Arkansas	0.003527337
California	0.888506151
Colorado	0.623092236
Connecticut	0.811544992
Illinois	0.896825397
Maine	0.473389356
Massachusetts	0.865828092
Minnesota	0.473948525
Missouri	0.680000000
Montana	0.708812261
New York	0.859822073
North Dakota	0.455555556
Ohio	0.491606715
Oregon	0.593052109
Pennsylvania	0.834966478
South Dakota	0.742627346
Texas	0.626387176
Utah	0.679933665
Vermont	0.683431953
Washington	0.326592518

A tibble: 21 × 2

Which state (of those that report `mmr`) has the smallest fraction of schools above the CDC target vaccination rate of 95%?

[12]:
```
low.state = df_vacc_rates |> slice_min(frac, n = 1)
low.state
```

state	frac
<chr>	<dbl>
Arkansas	0.003527337

A tibble: 1 × 2

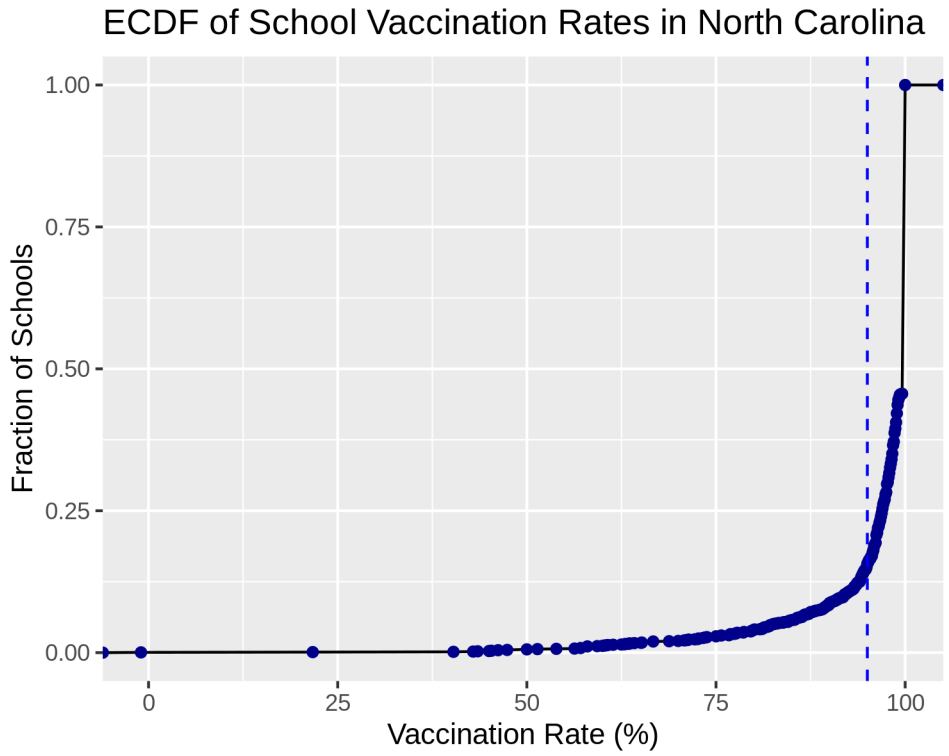
Make a ECDF plot of the school vaccination rates in North Carolina. Use the `overall` column which reports the "overall" vaccination rate at each school.

Remember, you can calculate the `y` value for ECDF charts using `row_number`, `mutate`, and `arrange`.

[33]:
```
nc_vacc_rates = us_vacc |>
    filter( state == 'North Carolina')
    #group_by(state == 'North Carolina')
#nc_vacc_rates
p = ggplot(nc_vacc_rates, aes(x = overall))
p = p + stat_ecdf(geom = "line") + stat_ecdf(geom = "point", color = 'darkblue')
p = p + labs(x = "Vaccination Rate (%)", y = "Fraction of Schools", title =␣
 ↪"ECDF of School Vaccination Rates in North Carolina")
p = p + geom_vline(xintercept = 95, linetype = "dashed", color = 'blue' )
#p = p + annotate("text", label = "CDC Recommended Vaccination Rate (95%)",
```

```
   # x = 70, y = 60, size = 2, colour = "red")

p
```



(My solution is below if you want to peek . Uncomment the code, change the cell to `markdown`, and exectute it.)

[ ]:
```
# ![](https://github.com/chuckpr/BIOS512/blob/main/assignments/measles-chart.
↪png?raw=true)
```

Is the median vaccination rate for NC schools above the CDC recommended target?

[57]:
```
nc_vacc_rates = us_vacc |>
    filter( state == 'North Carolina') |>
    select(overall) |>
    summarize(med_val = median(overall))
nc_vacc_rates
```

A tibble: 1×1

med_val
<dbl>
100

[ ]:
```

```