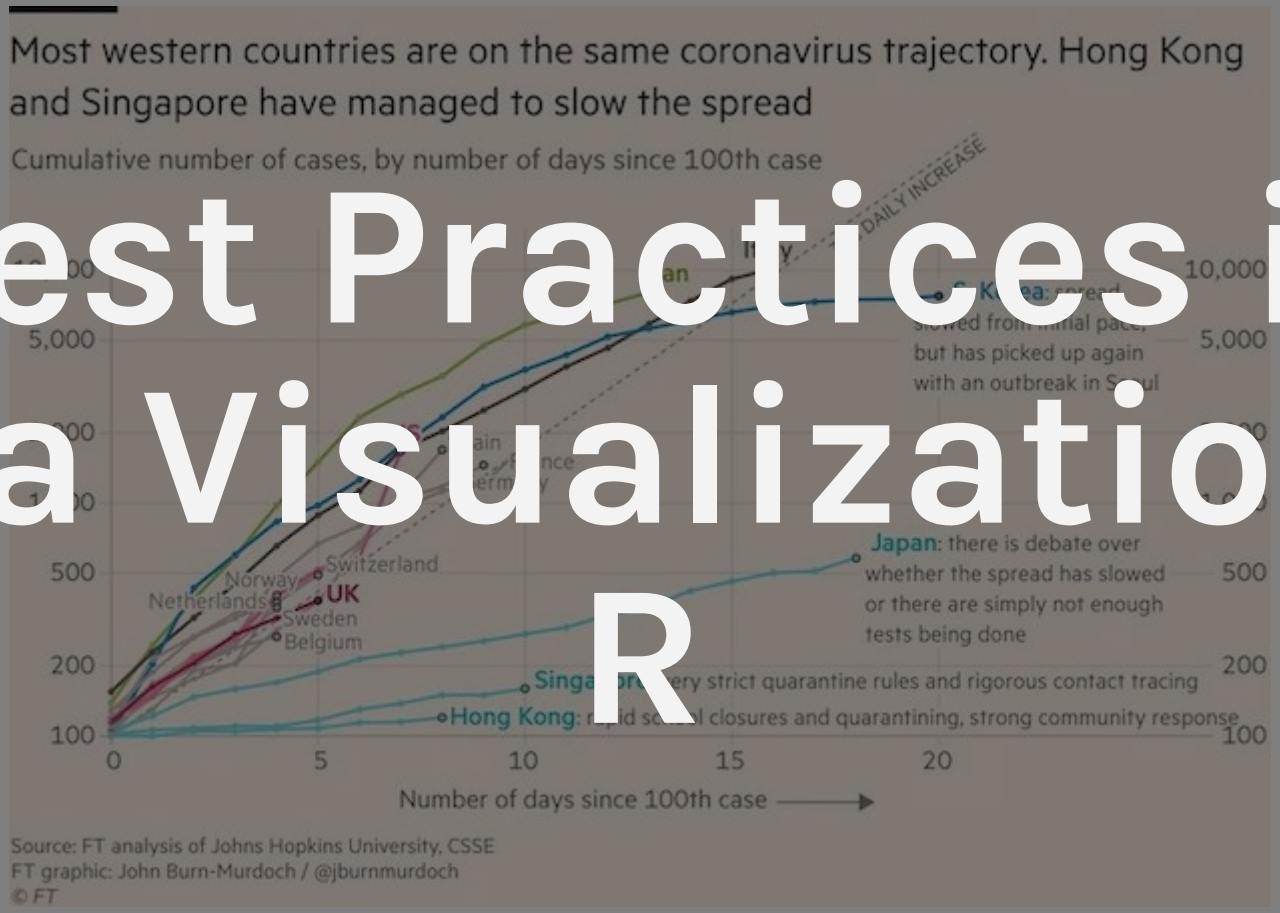
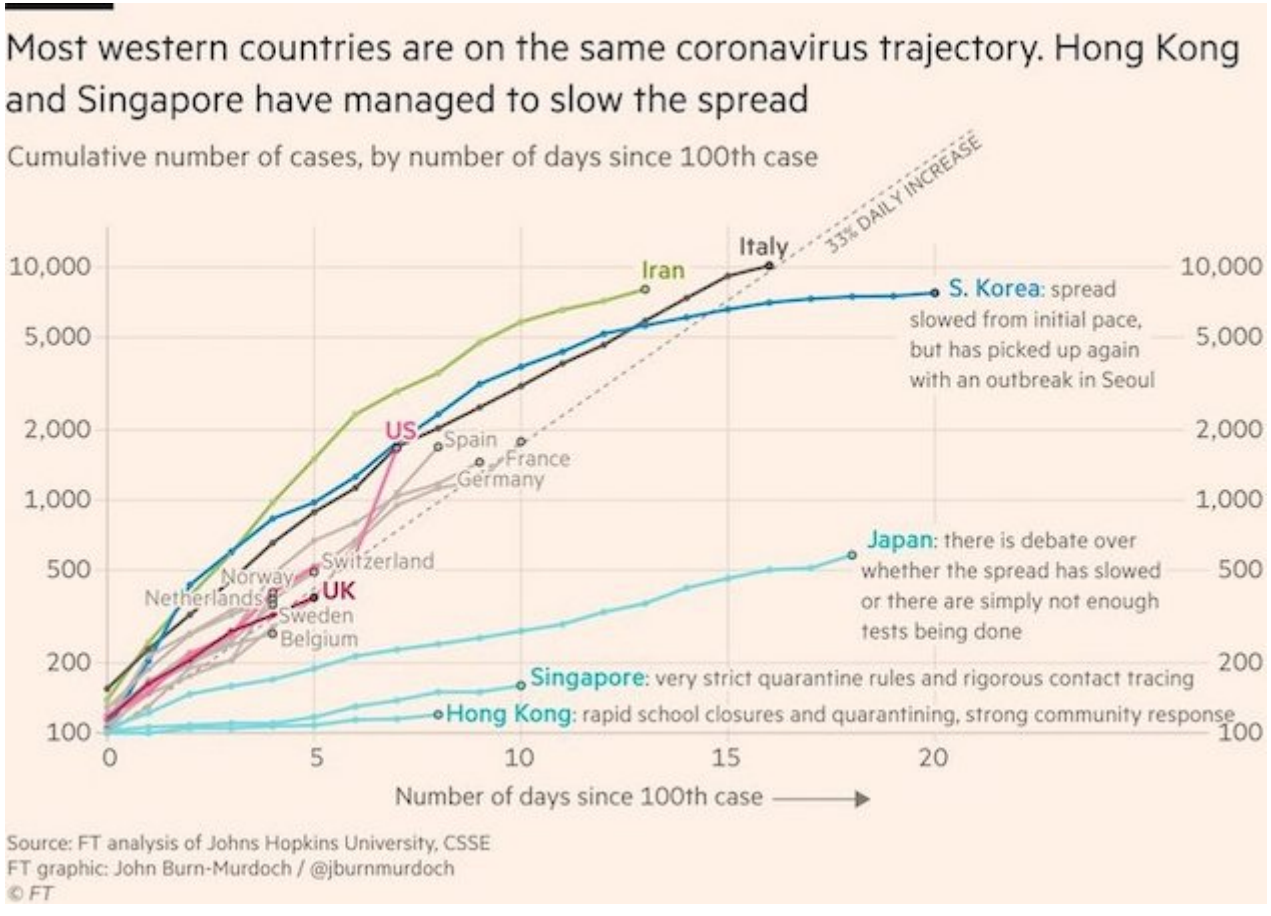
Advanced Data Visualization

Going Deeper with R

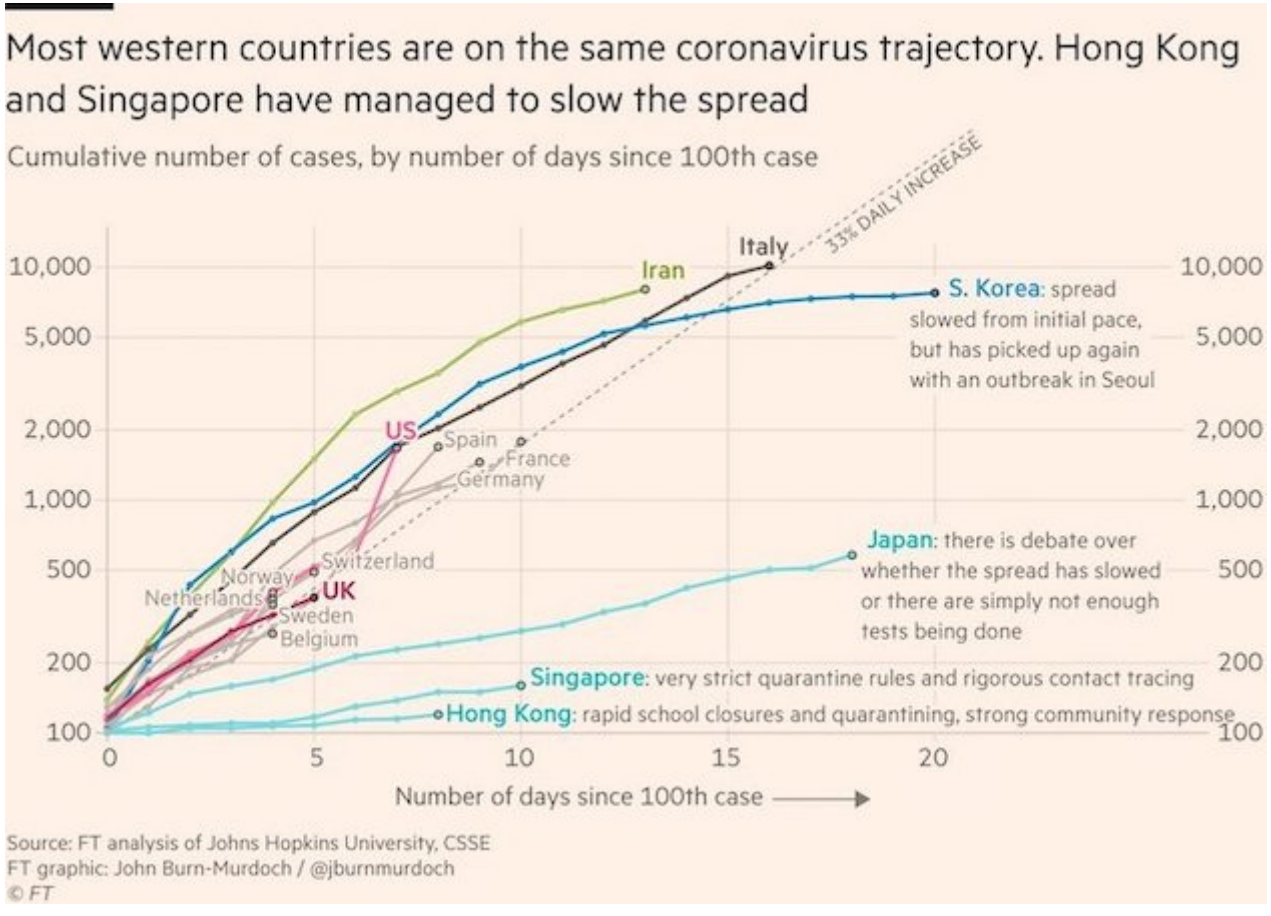


Best Practices in Data Visualization in R

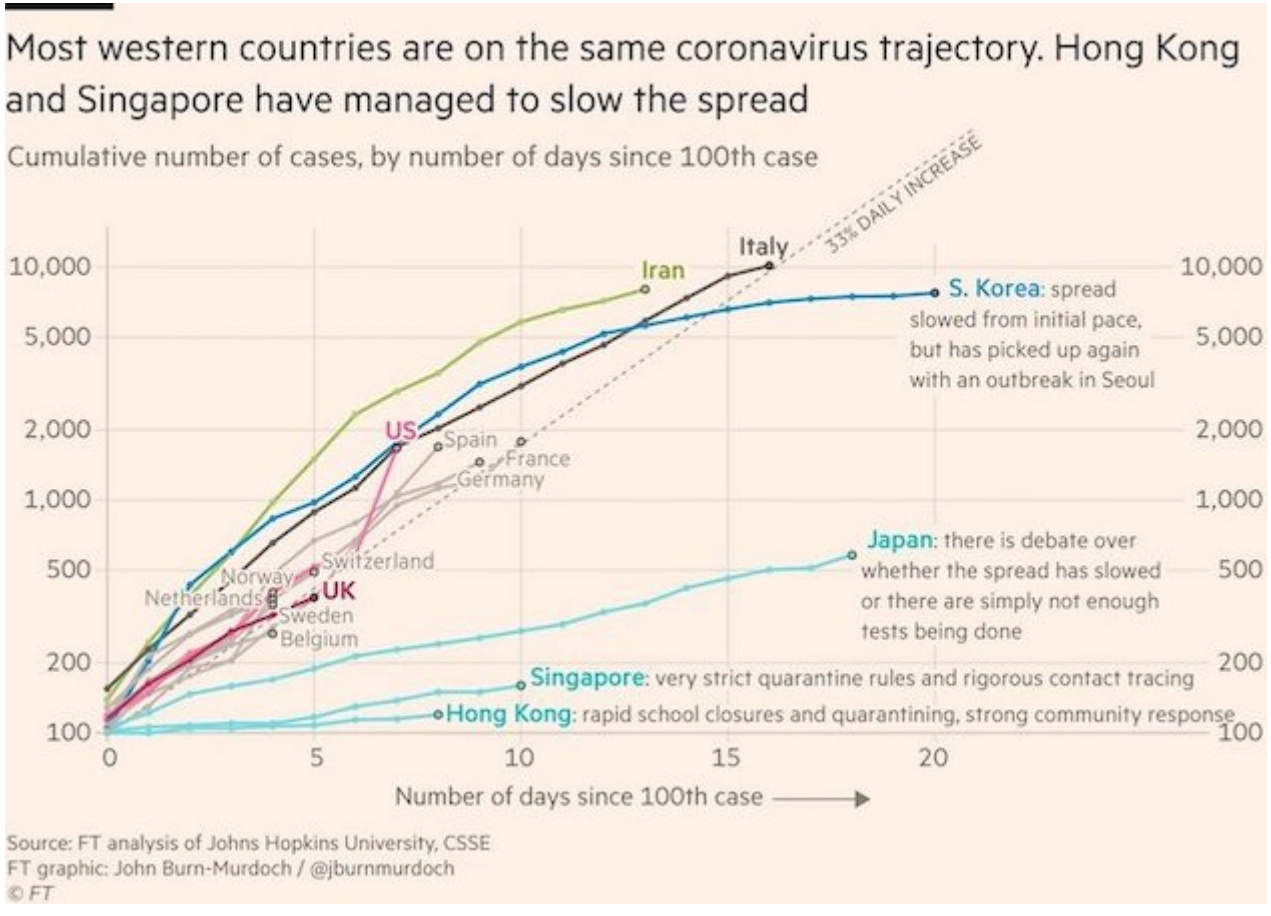


Source: [Financial Times, March 11, 2020]

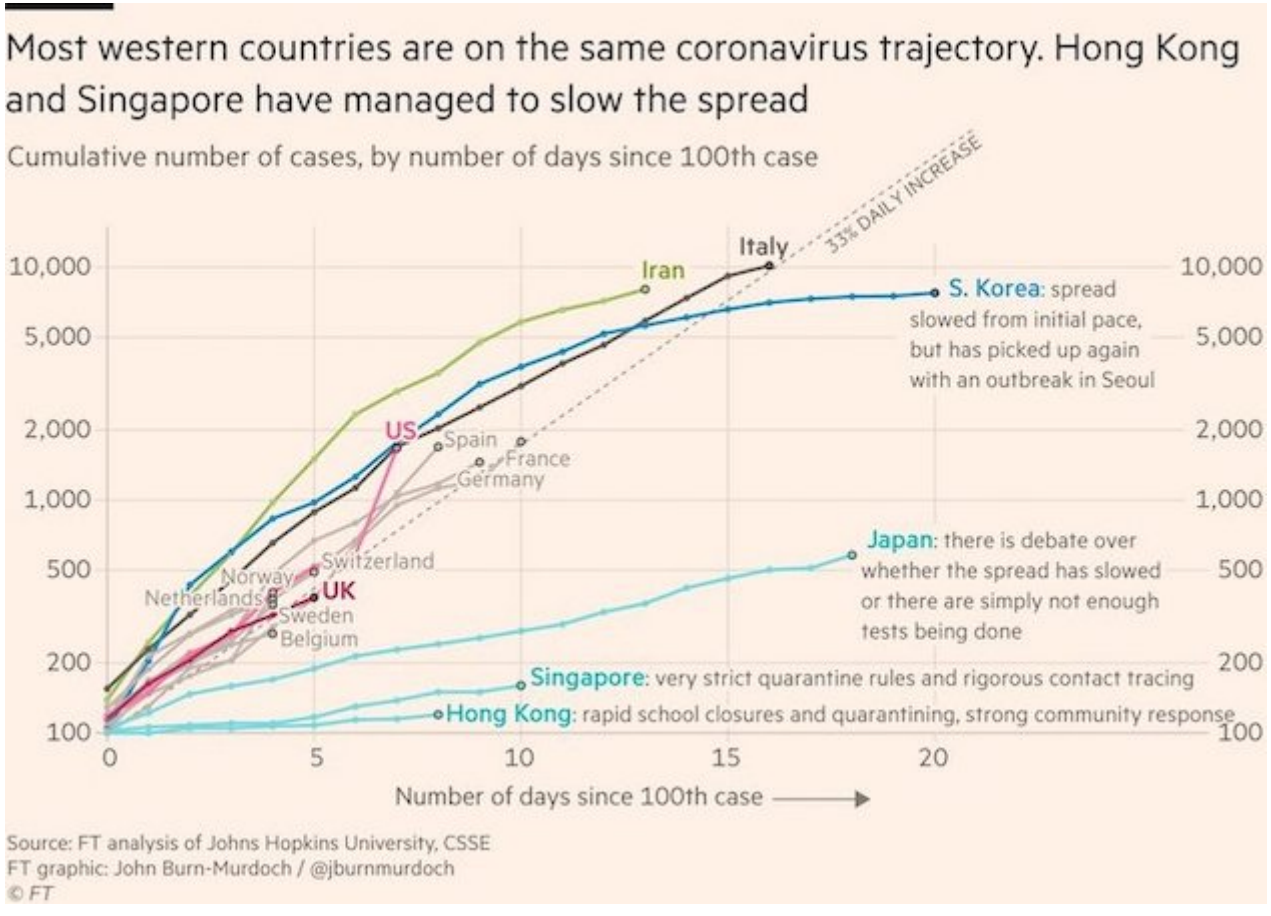

Highlight


Most western countries are on the same coronavirus trajectory. Hong Kong and Singapore have managed to slow the spread
Cumulative number of cases, by number of days since 100th case
33% DAILY INCREASE
Italy
Iran
S. Korea: spread slowed from initial pace, but has picked up again with an outbreak in Seoul
US
Spain
France
Germany
Switzerland
Norway
Netherlands
UK
Sweden
Belgium
Japan: there is debate over whether the spread has slowed or there are simply not enough tests being done
Singapore: very strict quarantine rules and rigorous contact tracing
Hong Kong: rapid school closures and quarantining, strong community response
10,000
5,000
2,000
1,000
500
200
100
0
5
10
15
20
Number of days since 100th case
Source: FT analysis of Johns Hopkins University, CSSE
FT graphic: John Burn-Murdoch / @jburnmurdoch
© FT

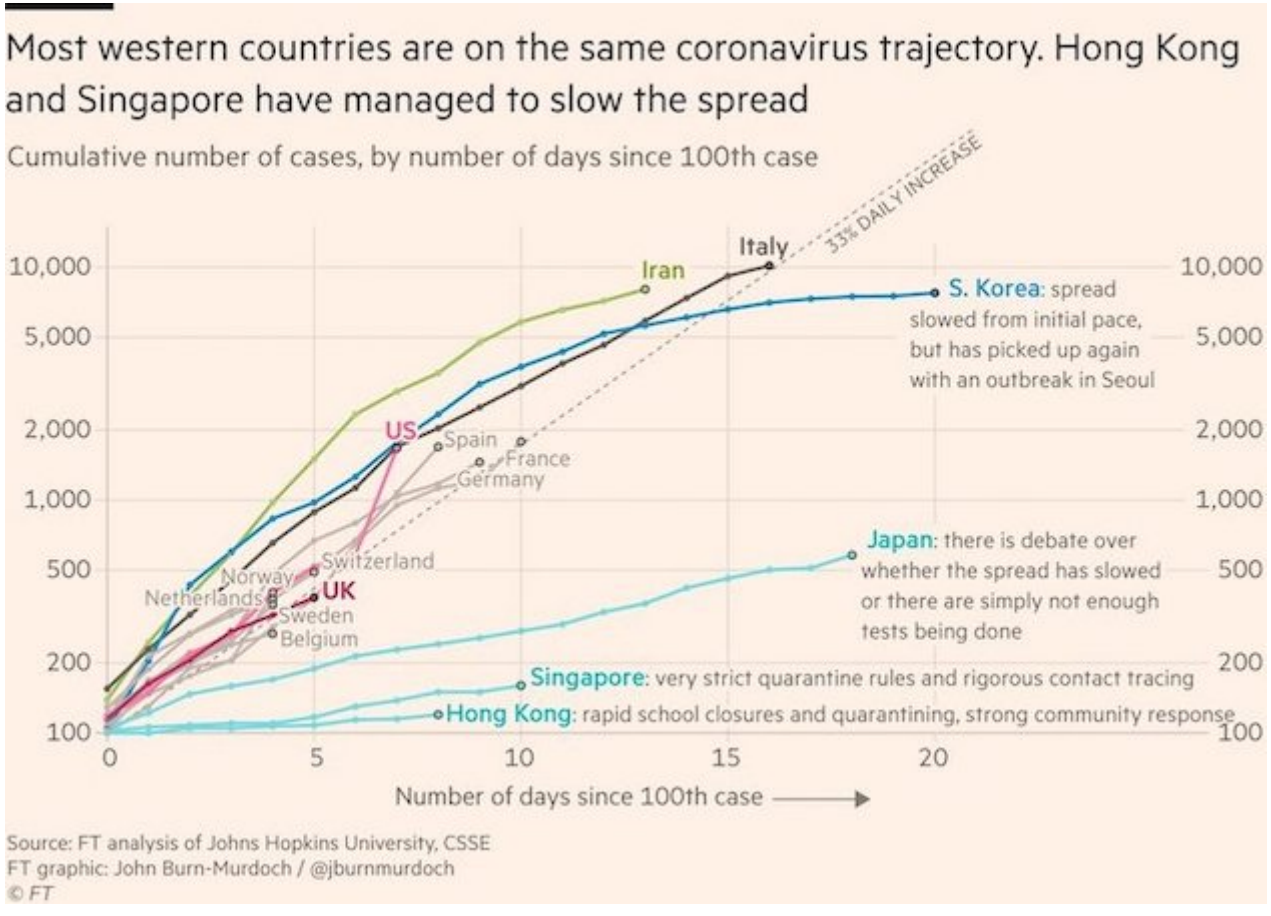
Declutter


Most western countries are on the same coronavirus trajectory. Hong Kong and Singapore have managed to slow the spread
Cumulative number of cases, by number of days since 100th case
33% DAILY INCREASE
Italy
Iran
S. Korea: spread slowed from initial pace, but has picked up again with an outbreak in Seoul
US
Spain
France
Germany
Switzerland
Norway
Netherlands
UK
Sweden
Belgium
Japan: there is debate over whether the spread has slowed or there are simply not enough tests being done
Singapore: very strict quarantine rules and rigorous contact tracing
Hong Kong: rapid school closures and quarantining, strong community response
10,000
5,000
2,000
1,000
500
200
100
0
5
10
15
20
Number of days since 100th case
Source: FT analysis of Johns Hopkins University, CSSE
FT graphic: John Burn-Murdoch / @jburnmurdoch
© FT

Explain


Most western countries are on the same coronavirus trajectory. Hong Kong and Singapore have managed to slow the spread
Cumulative number of cases, by number of days since 100th case
33% DAILY INCREASE
Italy
Iran
S. Korea: spread slowed from initial pace, but has picked up again with an outbreak in Seoul
US
Spain
France
Germany
Switzerland
Norway
Netherlands
UK
Sweden
Belgium
Japan: there is debate over whether the spread has slowed or there are simply not enough tests being done
Singapore: very strict quarantine rules and rigorous contact tracing
Hong Kong: rapid school closures and quarantining, strong community response
10,000
5,000
2,000
1,000
500
200
100
0
5
10
15
20
Number of days since 100th case
Source: FT analysis of Johns Hopkins University, CSSE
FT graphic: John Burn-Murdoch / @jburnmurdoch
© FT


Sparkle


Most western countries are on the same coronavirus trajectory. Hong Kong and Singapore have managed to slow the spread
Cumulative number of cases, by number of days since 100th case
33% DAILY INCREASE
Italy
Iran
S. Korea: spread slowed from initial pace, but has picked up again with an outbreak in Seoul
US
Spain
France
Germany
Switzerland
Norway
Netherlands
UK
Sweden
Belgium
Japan: there is debate over whether the spread has slowed or there are simply not enough tests being done
Singapore: very strict quarantine rules and rigorous contact tracing
Hong Kong: rapid school closures and quarantining, strong community response
10,000
5,000
2,000
1,000
500
200
100
0
5
10
15
20
Number of days since 100th case
Source: FT analysis of Johns Hopkins University, CSSE
FT graphic: John Burn-Murdoch / @jburnmurdoch
© FT


Tidy Data

Can We Plot Untidy Data?

```
german_speakers_numeric <- read_excel(path = "data-raw/german-and-french-speakers.xlsx",
                                      sheet = "German speakers",
                                      na = "-") %>%
  clean_names()
```

```
german_speakers_numeric
```

state	number_of_germ...	number_of_germ...	number_of_germ...
<chr>	<dbl>	<dbl>	<dbl>
Alabama	426	395	711
Alaska	331	201	131
Arizona	636	858	136
Arkansas	NA	635	557
California	440	318	854

1-5 of 51 rows

Previous **1** 2 3 4 5 6 … 11 Next

Can We Plot Untidy Data?

```
ggplot(data = german_speakers_numeric,
       mapping = aes(x = ???,
                     y = state))
```

Can We Plot Untidy Data?

```
german_speakers_tidy <- german_speakers_numeric %>%
  pivot_longer(cols = -state,
               names_to = "year",
               values_to = "number") %>%
  mutate(year = parse_number(year))
```

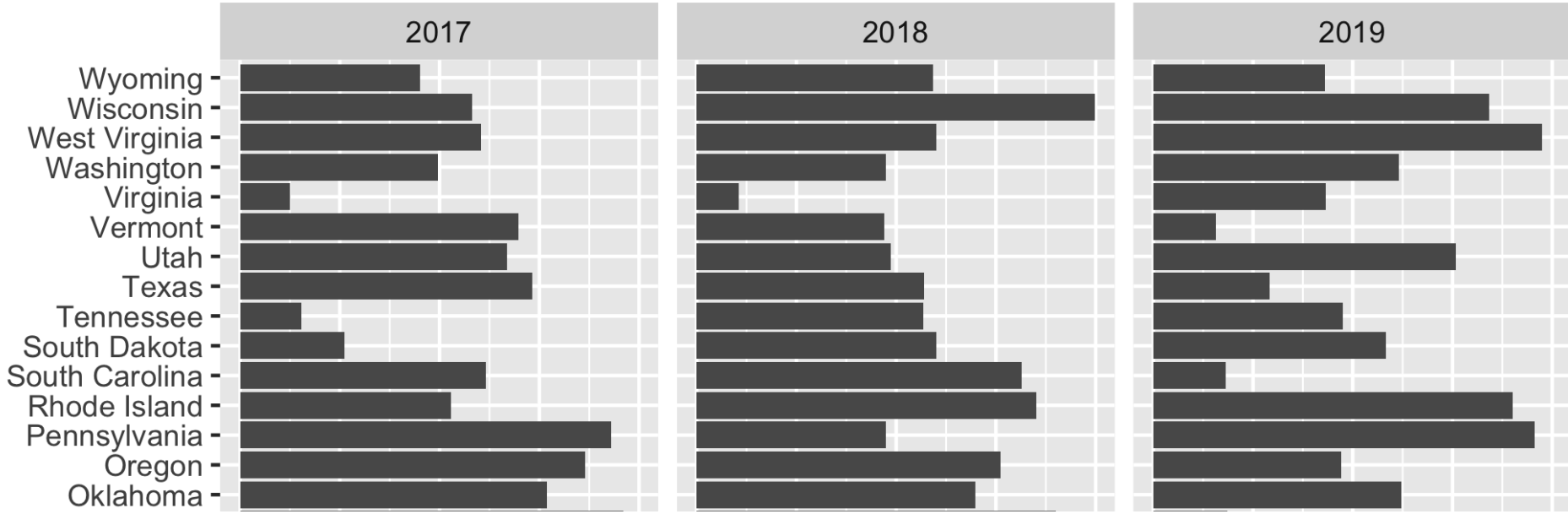
```
german_speakers_tidy
```

state <chr>	year <dbl>	number <dbl>
Alabama	2017	426
Alabama	2018	395
Alabama	2019	711
Alaska	2017	331
Alaska	2018	201

1-5 of 153 rows Previous **1** 2 3 4 5 6 31 Next

Can We Plot Untidy Data?

```
ggplot(data = german_speakers_tidy,
       mapping = aes(x = number,
                     y = state)) +
  geom_col() +
  facet_wrap(~year)
```



Pipe Data Into ggplot

Load Data

```
third_grade_math_proficiency <- read_rds("data/third_grade_math_proficiency.rds")
```

```
third_grade_math_proficiency
```

school <chr>	school_id <dbl>	district <chr>	district_id <dbl>	percent_profi... <dbl>
Brooklyn Primary School	2	Baker SD 5J	1894	0.44954128
Haines Elementary School	4	Baker SD 5J	1894	0.86363636
Pine Eagle Charter School	15	Pine Eagle SD 61	1897	0.58333333
Alsea Charter School	17	Alsea SD 7J	1899	0.33333333

Pipe Data Into ggplot

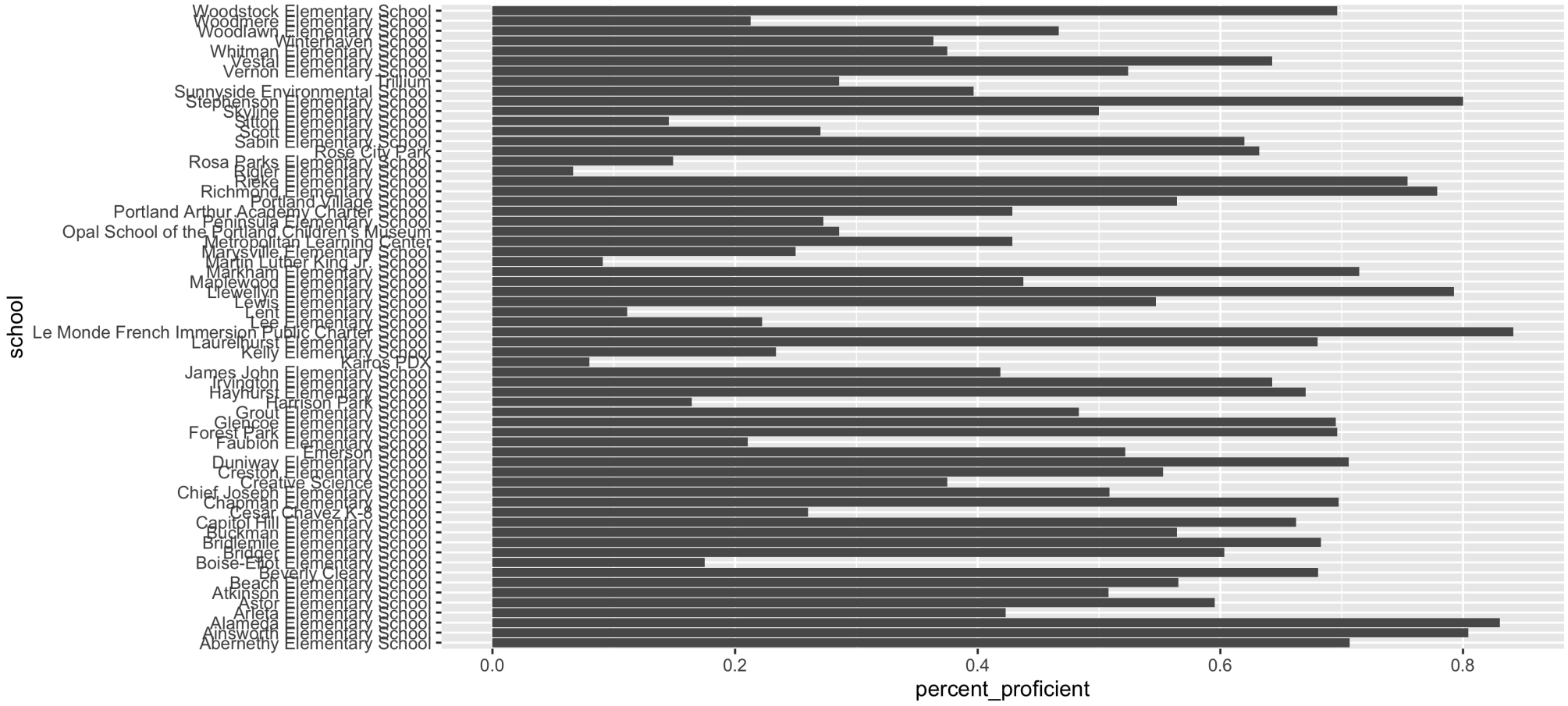
```
third_grade_math_proficiency %>%
  filter(year == "2018-2019") %>%
  filter(district == "Portland SD 1J")
```

school <chr>	school_id <dbl>	district <chr>	district_id <dbl>	percent_profi… <dbl>
Abernethy Elementary School	822	Portland SD 1J	2180	0.70652174
Ainsworth Elementary School	823	Portland SD 1J	2180	0.80434783
Alameda Elementary School	824	Portland SD 1J	2180	0.83050847

Pipe Data Into ggplot

```
third_grade_math_proficiency %>%
  filter(year == "2018-2019") %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = school, y = percent_proficient)) +
  geom_col() +
  coord_flip()
```

Pipe Data Into ggplot



Your Turn

1. Create a new RMarkdown document

2. Create a data frame called `enrollment_by_race_ethnicity` by reading in your race/ethnicity data from the data wrangling and analysis section using the `read_rds()` function

3. Pipe your data into a bar chart that shows the breakdown of race/ethnicity among students in Beaverton SD 48J in 2018-2019

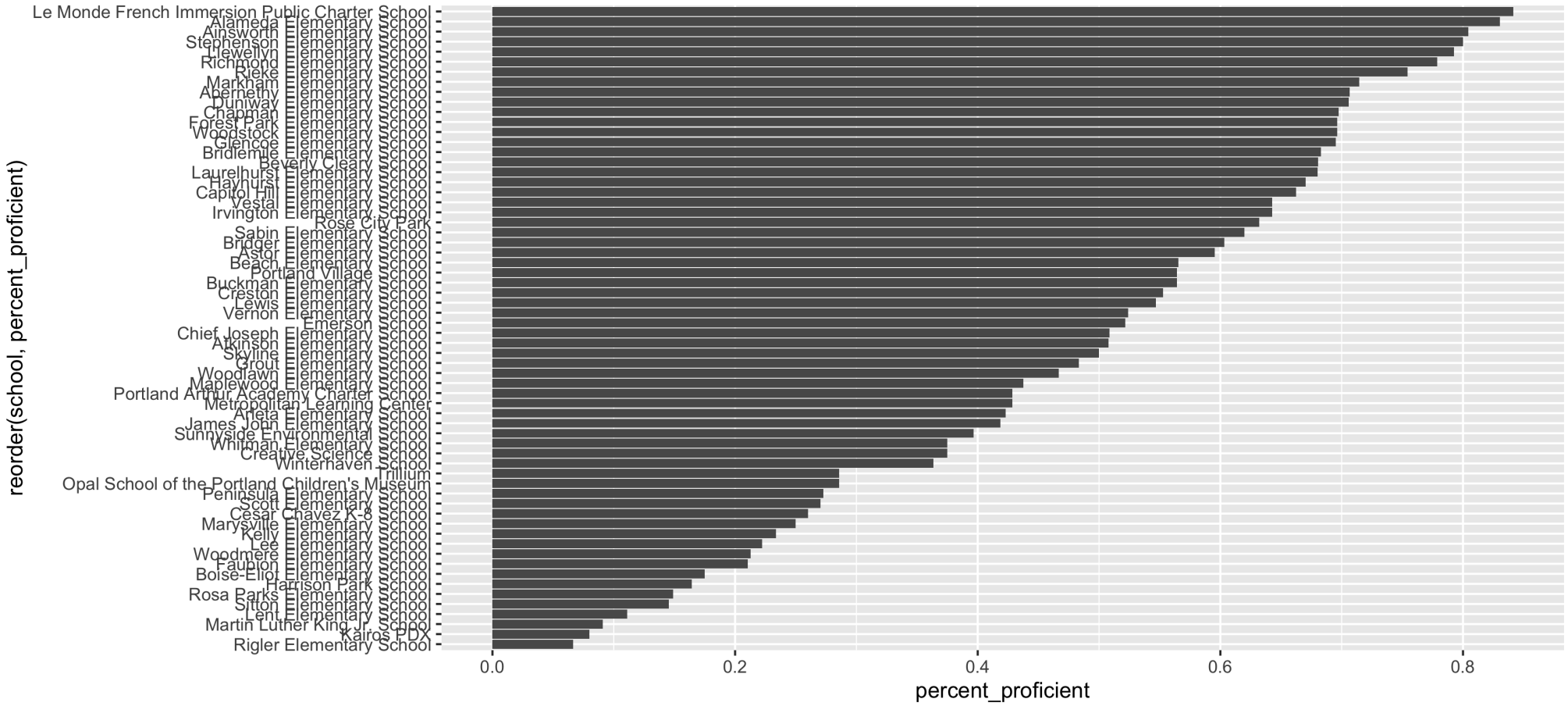

Highlight

Reorder Plots to Highlight Findings

Reorder Plots to Highlight Findings

```
third_grade_math_proficiency %>%
  filter(year == "2018-2019") %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = reorder(school, percent_proficient),
             y = percent_proficient)) +
  geom_col() +
  coord_flip()
```

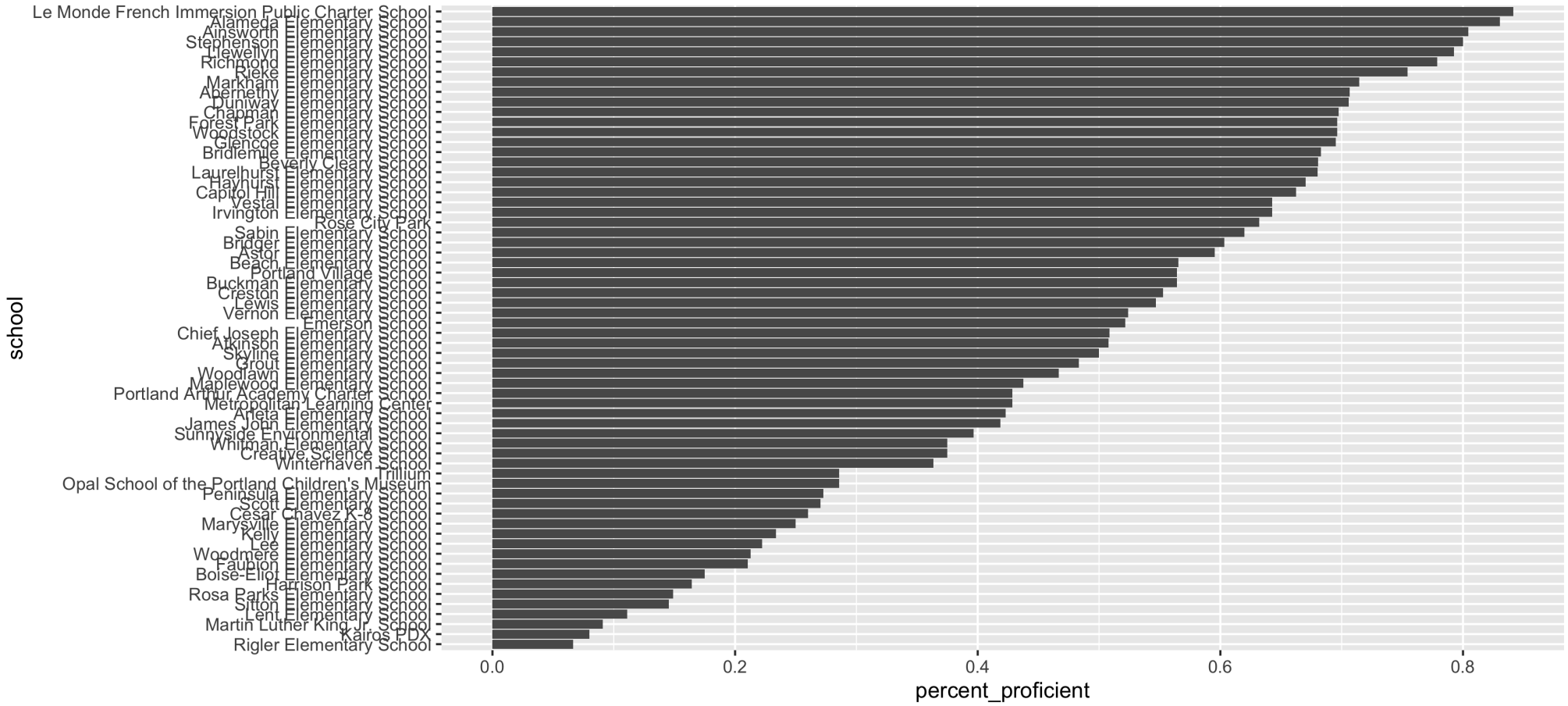
Reorder Plots to Highlight Findings



Reorder Plots to Highlight Findings

```
third_grade_math_proficiency %>%
  filter(year == "2018-2019") %>%
  filter(district == "Portland SD 1J") %>%
  mutate(school = fct_reorder(school, percent_proficient)) %>%
  ggplot(aes(x = school,
             y = percent_proficient)) +
  geom_col() +
  coord_flip()
```

Reorder Plots to Highlight Findings



Your Turn

Make a bar chart that shows race/ethnicity in Beaverton SD 48J. As before, filter your data to only include 2018-2019 data and only include Beaverton SD 48J. Then, do the following:

1. Using the `reorder()` function, make a bar chart that shows the percent of race/ethnicity groups in descending order

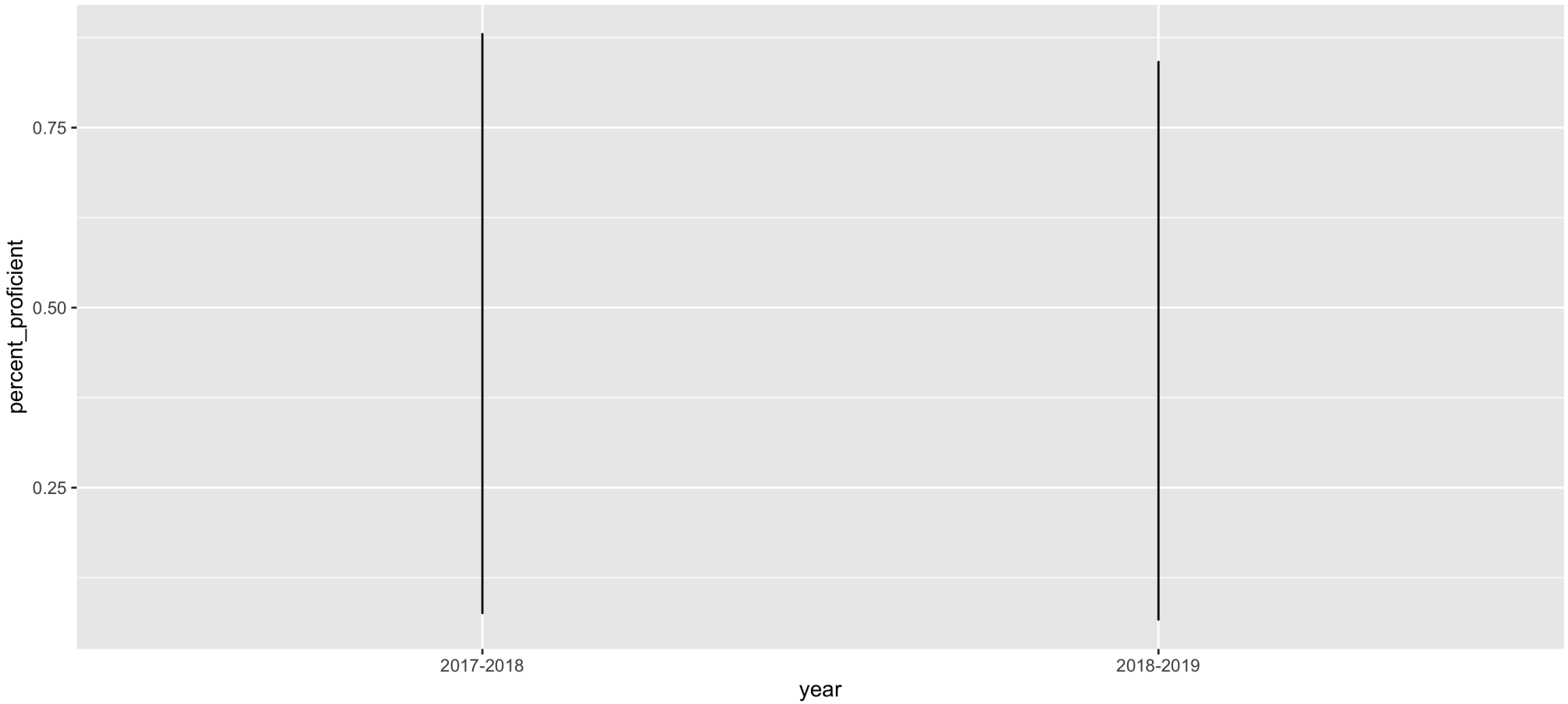
2. Make the same bar chart using `mutate()` and `fct_reorder()` to reorder the race/ethnicity groups

Line Charts

Line Charts

```
third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year,
             y = percent_proficient)) +
  geom_line()
```

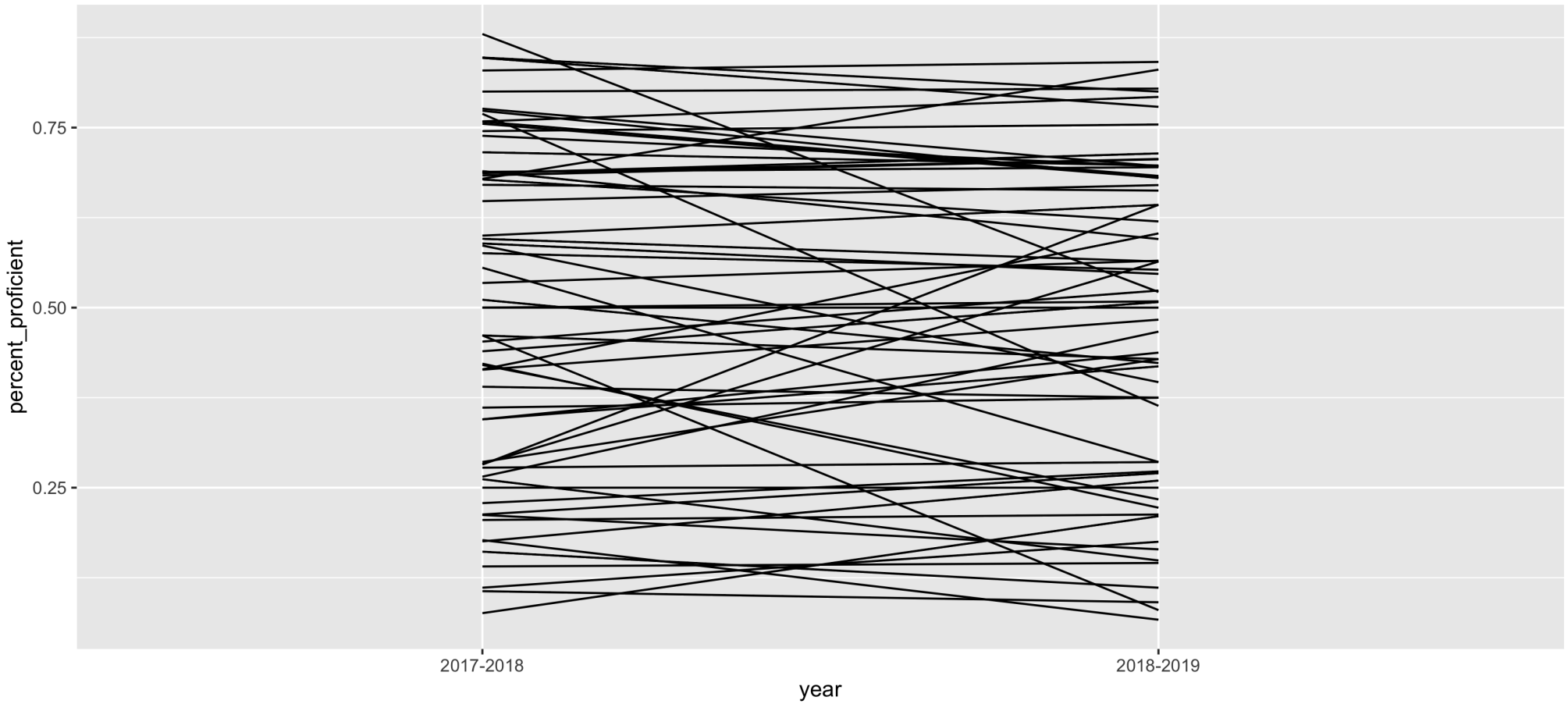
Line Charts



Line Charts

```
third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year,
             y = percent_proficient,
             group = school)) +
  geom_line()
```

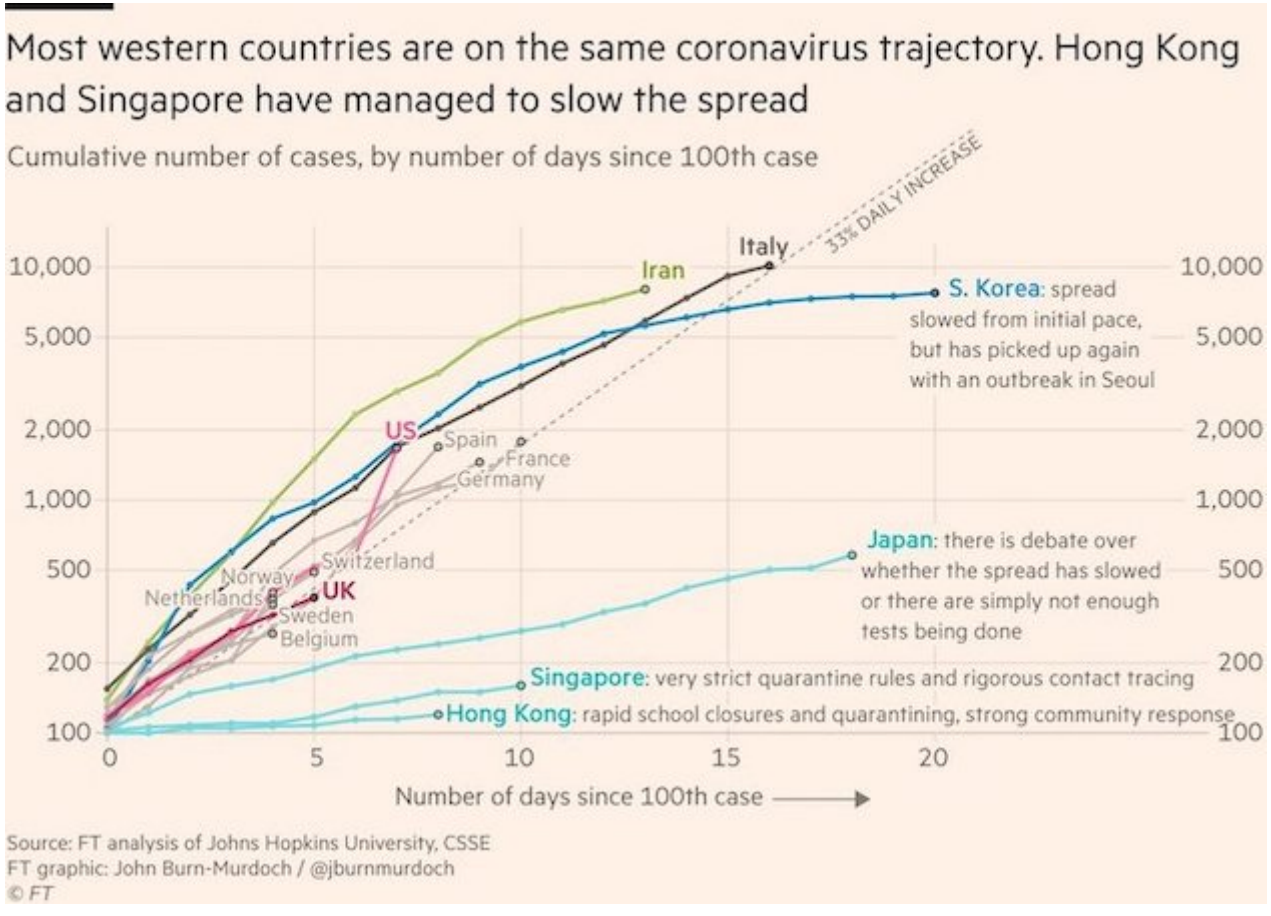
Line Charts



Your Turn

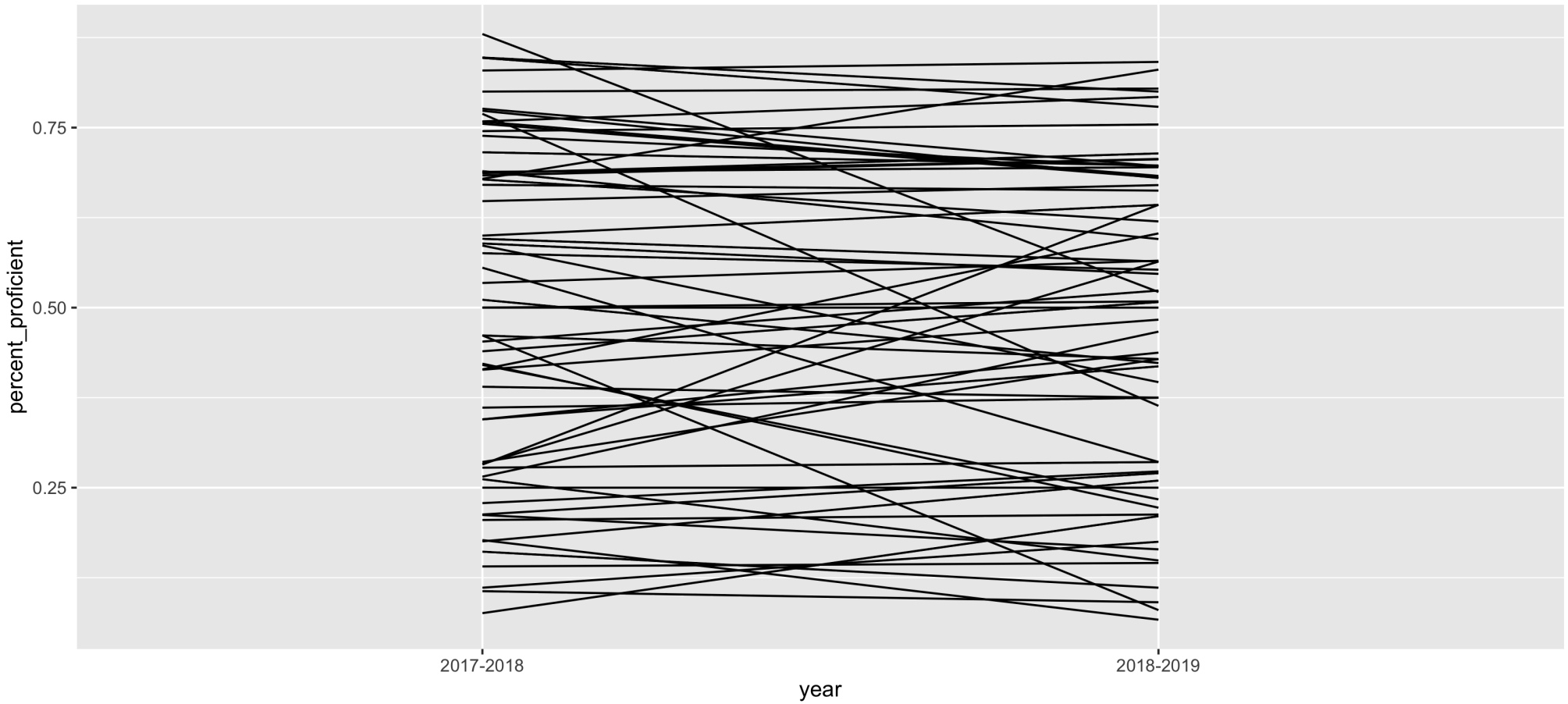
Make a line chart that shows the growth in the Hispanic/Latino population in school districts from 2017-2018 to 2018-2019

Use Color to Highlight Findings



Source: [Financial Times, March 11, 2020]

Use Color to Highlight Findings



Use Color to Highlight Findings

```
highlight_school <- third_grade_math_proficiency %>%
  filter(school == "Vestal Elementary School")
```

```
highlight_school
```

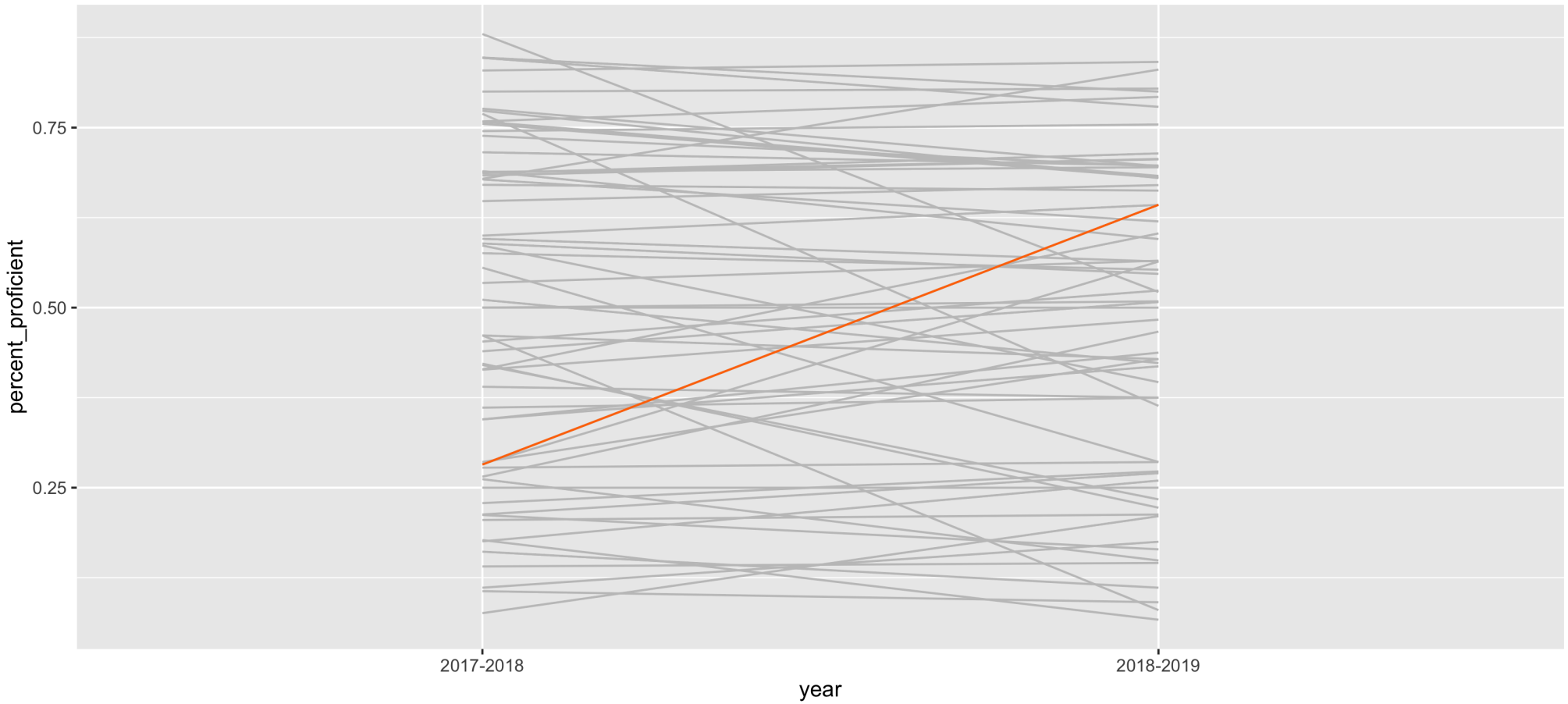
school <chr>	school_id <dbl>	district <chr>	district_id <dbl>	percent_profi... <dbl>
Vestal Elementary School	896	Portland SD 1J	2180	0.2820513
Vestal Elementary School	896	Portland SD 1J	2180	0.6428571

2 rows | 1-5 of 6 columns

Use Color to Highlight Findings

```
rru_orange <- "#FF7400"
rru_gray <- "#C3C3C3"

third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange)
```

Use Color to Highlight Findings



Your Turn

1. Identify one school district that has had a lot of growth in its Hispanic/Latino population from 2017-2018 to 2018-2019

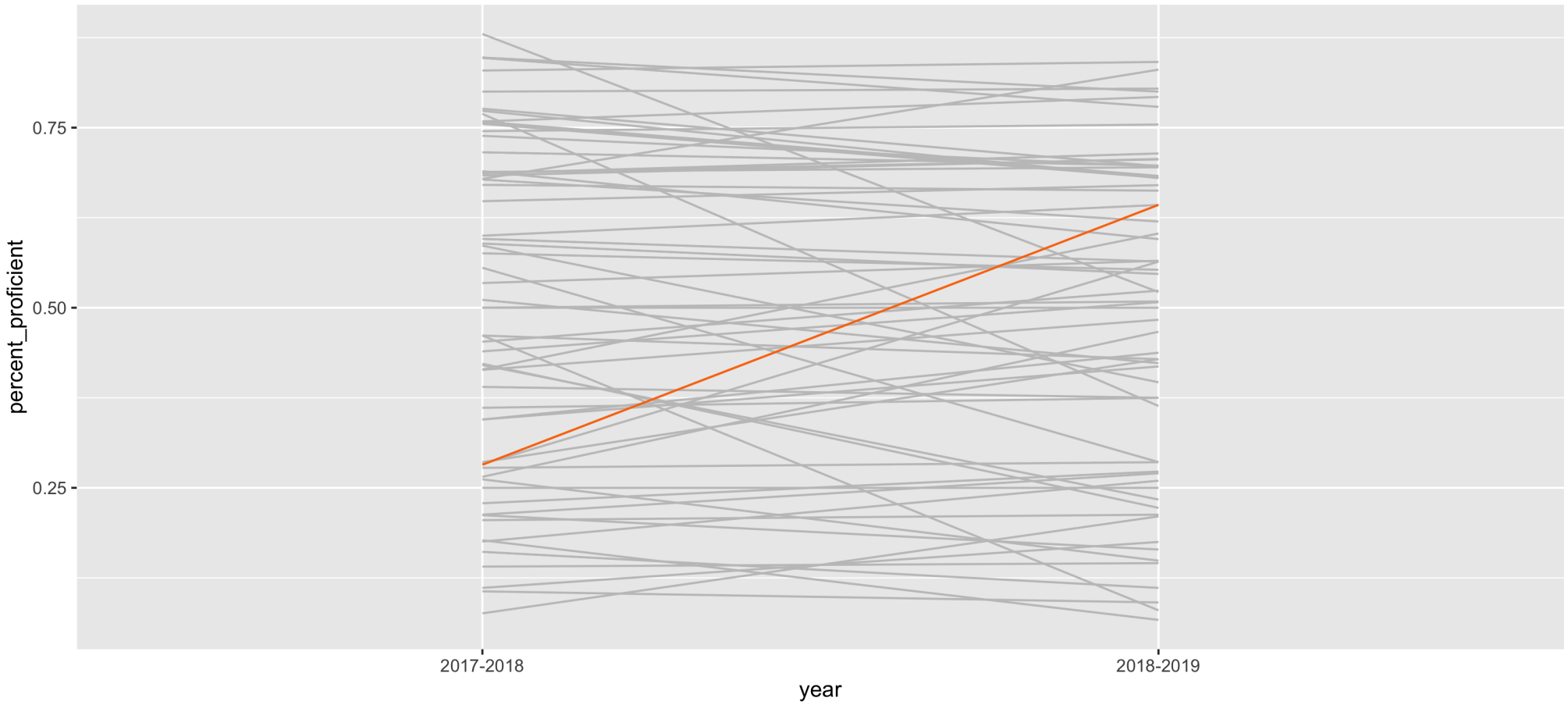
2. Create a new data frame called `highlight_district` and only include this district in it

3. Use the `highlight_district` data frame to create a new `geom_line()` layer on top of the other data

4. Make sure this new layer is a bright color and all other layers are some type of light gray

Declutter

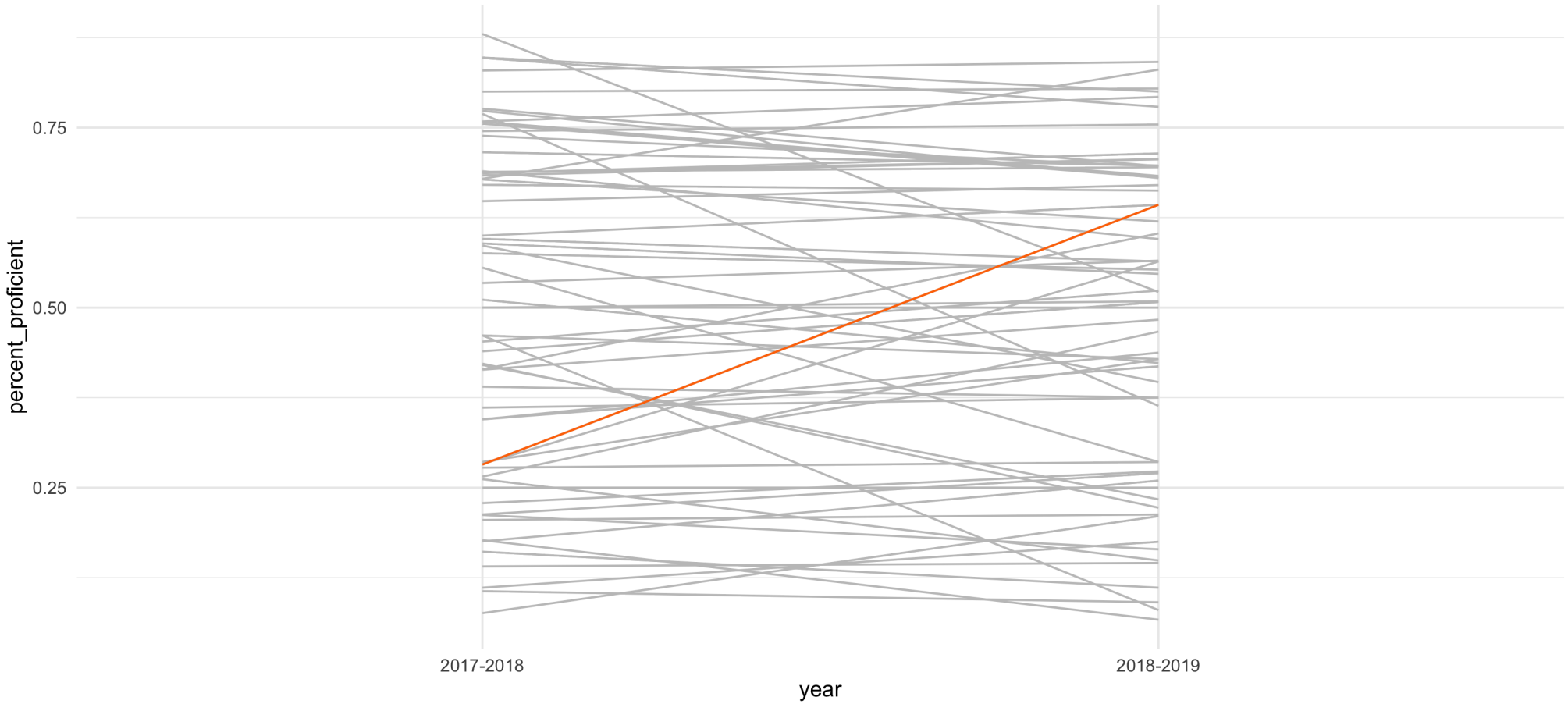
Remove the Default Gray Background



Remove the Default Gray Background

```
third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  theme_minimal()
```

Remove the Default Gray Background



Remove Axis Titles

Modify components of a theme

Source: `R/theme.r`

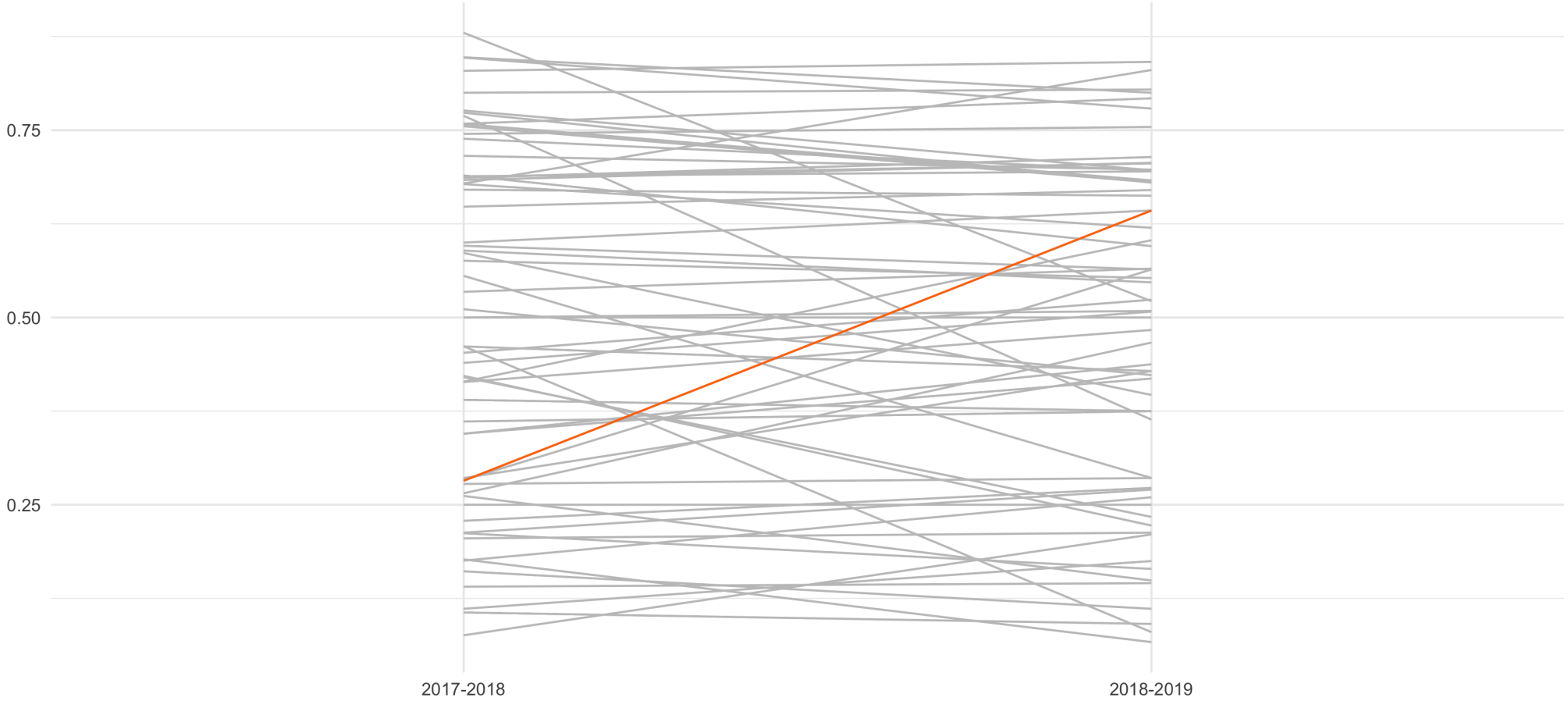
Themes are a powerful way to customize the non-data components of your plots: i.e. titles, labels, fonts, background, gridlines, and legends. Themes can be used to give plots a consistent customized look. Modify a single plot's theme using `theme()`; see `theme_update()` if you want modify the active theme, to affect all subsequent plots. Use the themes available in complete themes if you would like to use a complete theme such as `theme_bw()`, `theme_minimal()`, and more. Theme elements are documented together according to inheritance, read more about theme inheritance below.

Source: `ggplot2` documentation

Remove Axis Titles

```
third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  theme_minimal() +
  theme(axis.title = element_blank())
```

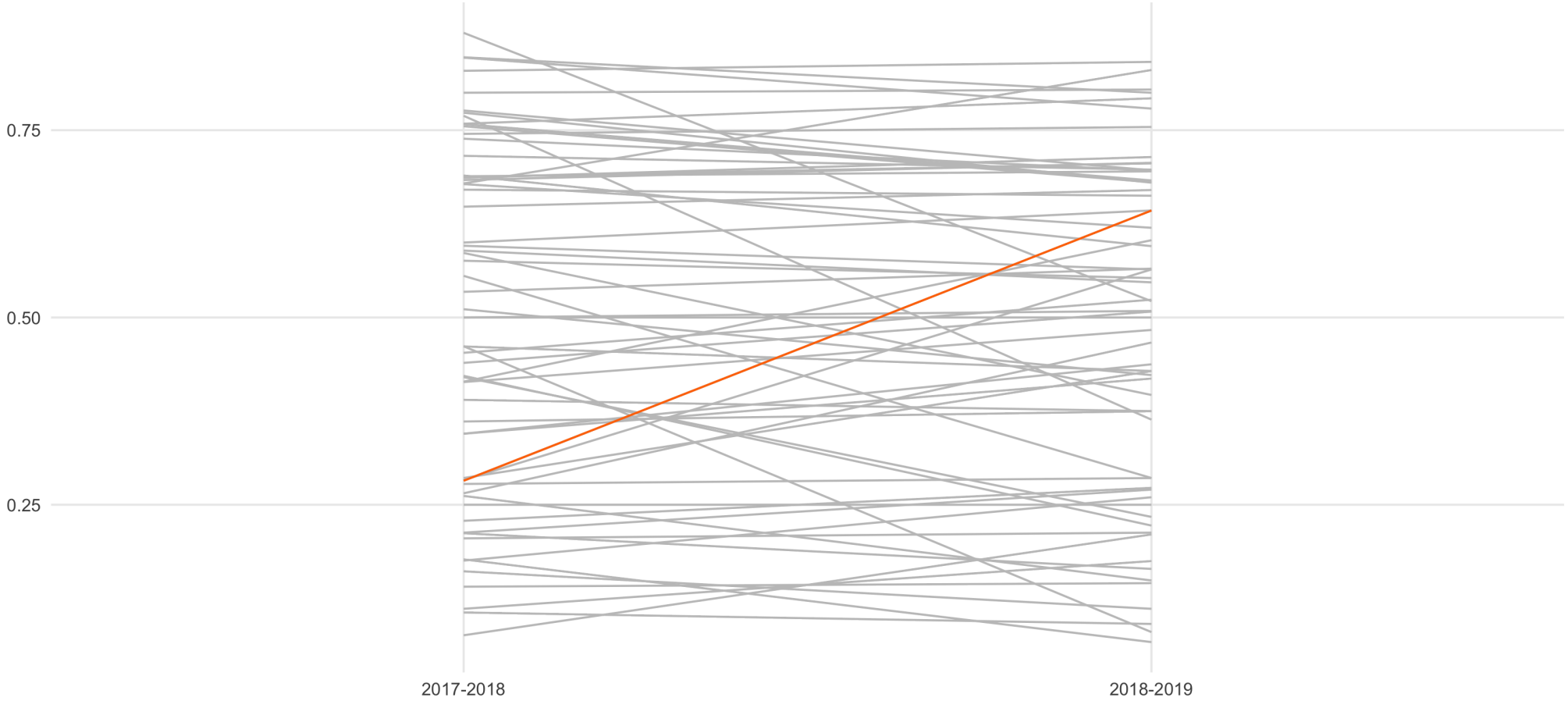
Remove Axis Titles



Remove or Minimize Grid Lines

```
third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  theme_minimal() +
  theme(axis.title = element_blank(),
        panel.grid.minor = element_blank())
```

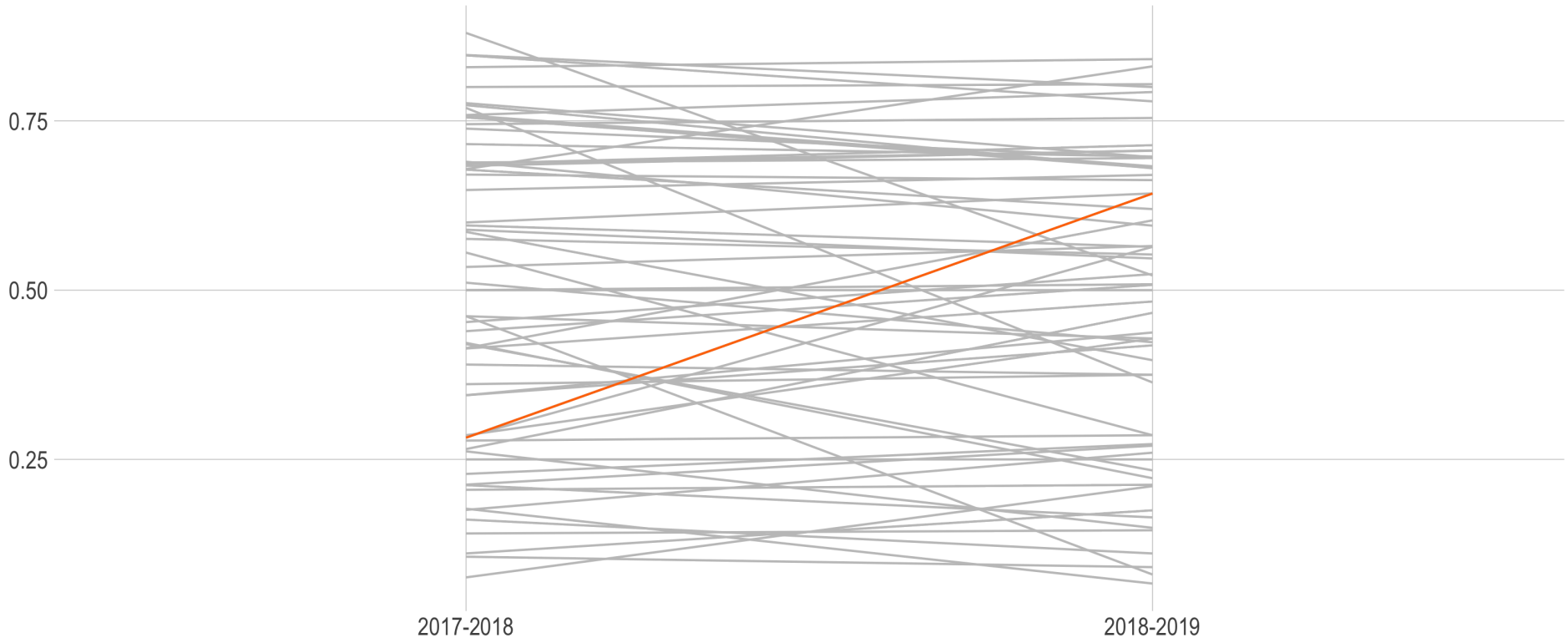
Remove or Minimize Grid Lines



Work with Existing Themes

```
library(hrbrthemes)

third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  theme_ipsum(axis_title_size = 0,
              grid = "XY")
```

Work with Existing Themes



Your Turn

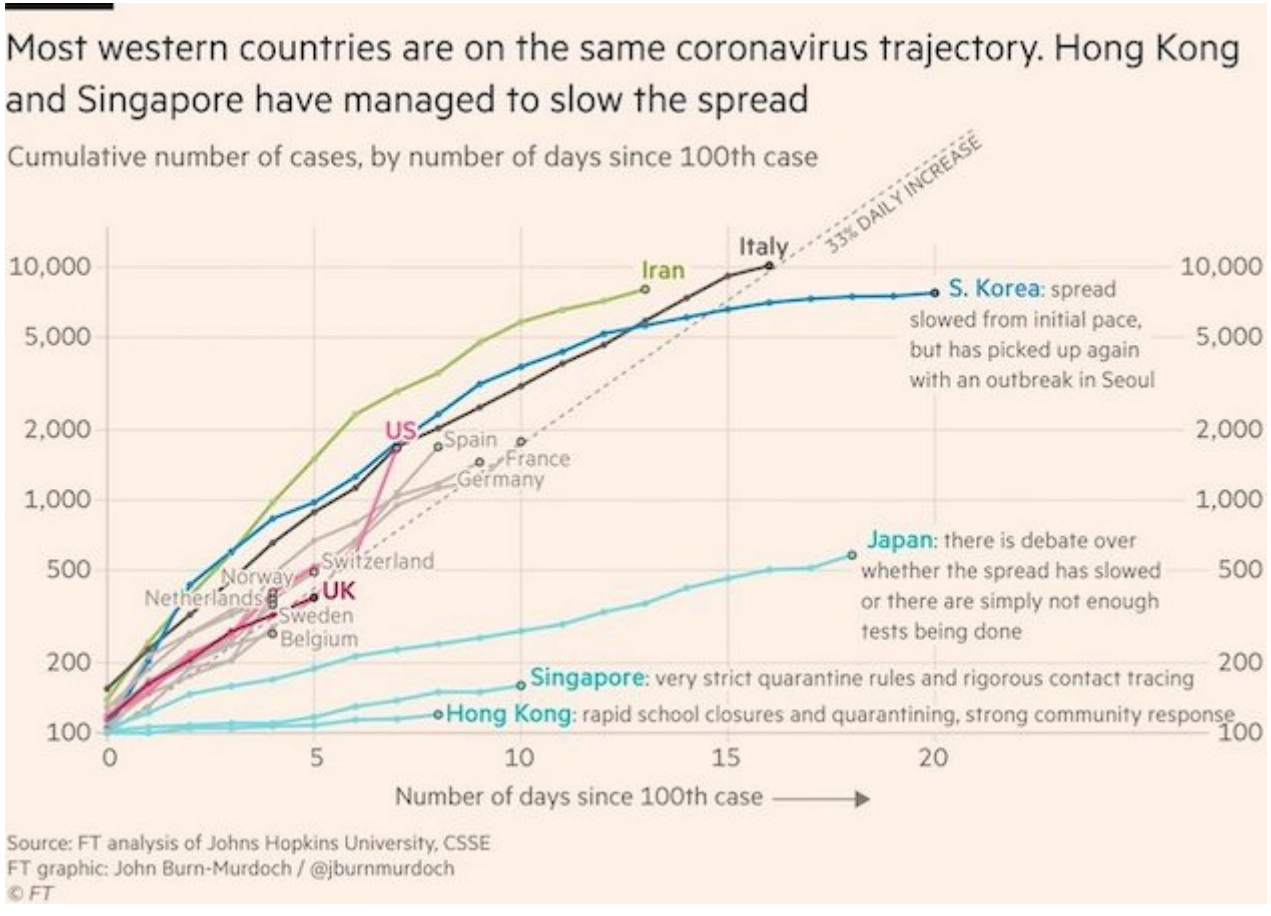
Use some combination of:

- Complete themes: e.g. `theme_minimal()` or `theme_ipsum()` from `hrbrthemes`
- The `theme()` function

Do the following:

1. Remove gray background
2. Remove axis titles
3. Remove or minimize grid lines

Explain


Most western countries are on the same coronavirus trajectory. Hong Kong and Singapore have managed to slow the spread
Cumulative number of cases, by number of days since 100th case
33% DAILY INCREASE
Italy
Iran
S. Korea: spread slowed from initial pace, but has picked up again with an outbreak in Seoul
US
Spain
France
Germany
Switzerland
Norway
Netherlands
UK
Sweden
Belgium
Japan: there is debate over whether the spread has slowed or there are simply not enough tests being done
Singapore: very strict quarantine rules and rigorous contact tracing
Hong Kong: rapid school closures and quarantining, strong community response
10,000
5,000
2,000
1,000
500
200
100
0
5
10
15
20
Number of days since 100th case
Source: FT analysis of Johns Hopkins University, CSSE
FT graphic: John Burn-Murdoch / @jburnmurdoch
© FT


Use the scales Package for
Nicely Formatted Values

scales

```
highlight_school
```

school <chr>	school_id <dbl>	district <chr>	district_id <dbl>	percent_profi... <dbl>
Vestal Elementary School	896	Portland SD 1J	2180	0.2820513
Vestal Elementary School	896	Portland SD 1J	2180	0.6428571

2 rows | 1-5 of 6 columns

scales

```
library(scales)

highlight_school <- third_grade_math_proficiency %>%
  filter(school == "Vestal Elementary School") %>%
  mutate(percent_proficient_display = percent(percent_proficient, accuracy = 1))
```

```
highlight_school %>%
  select(school, percent_proficient, percent_proficient_display)
```

school	percent_proficient	percent_proficient_disp…
<chr>	<dbl>	<chr>
Vestal Elementary School	0.2820513	28%
Vestal Elementary School	0.6428571	64%

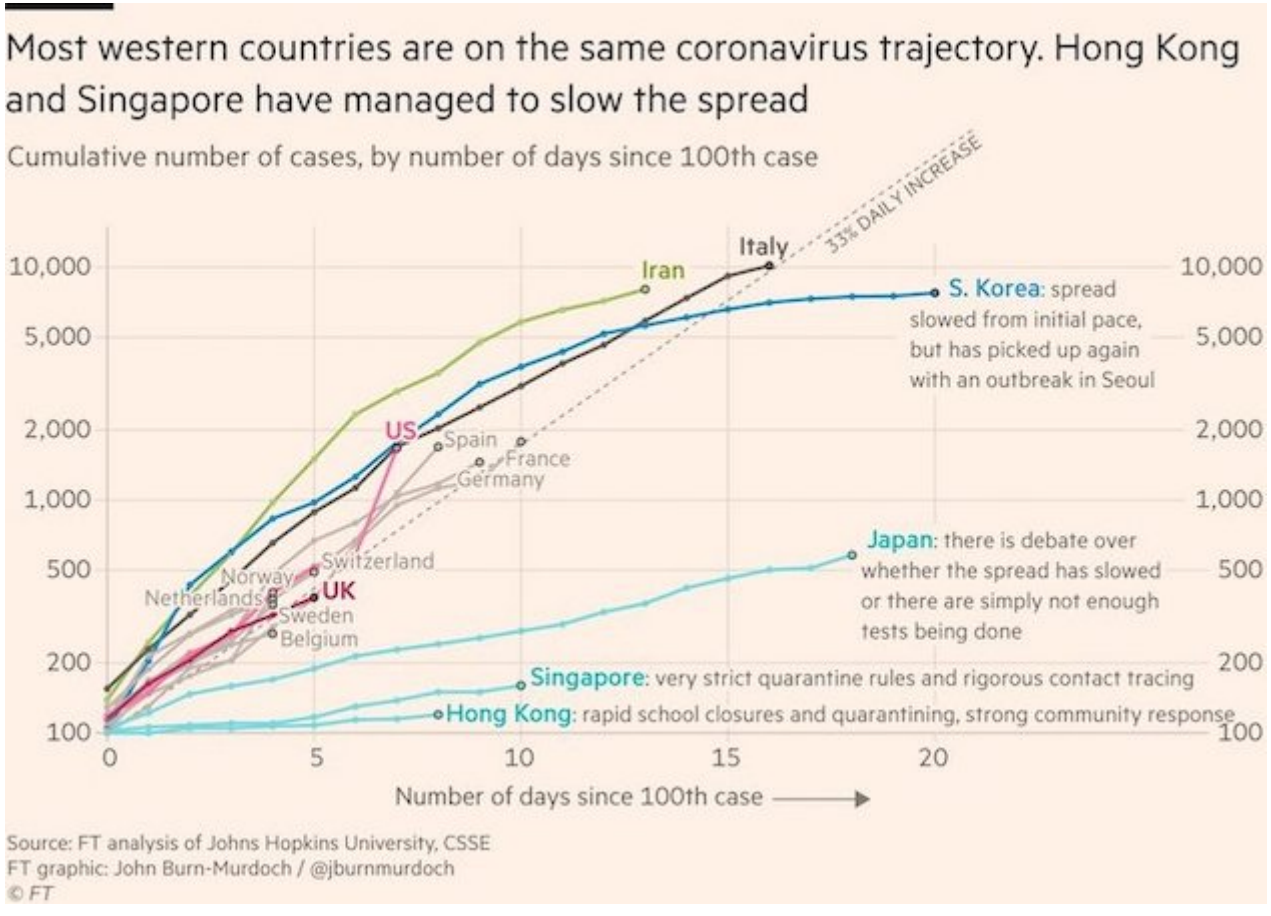
2 rows

Your Turn

1. Make a new variable called `percent_display` that shows the `percent_of_total_enrollment` variable as a nicely formatted percent (rounded to the nearest whole number)

2. Make sure you save this as `highlight_district` (i.e. don't just display the result)

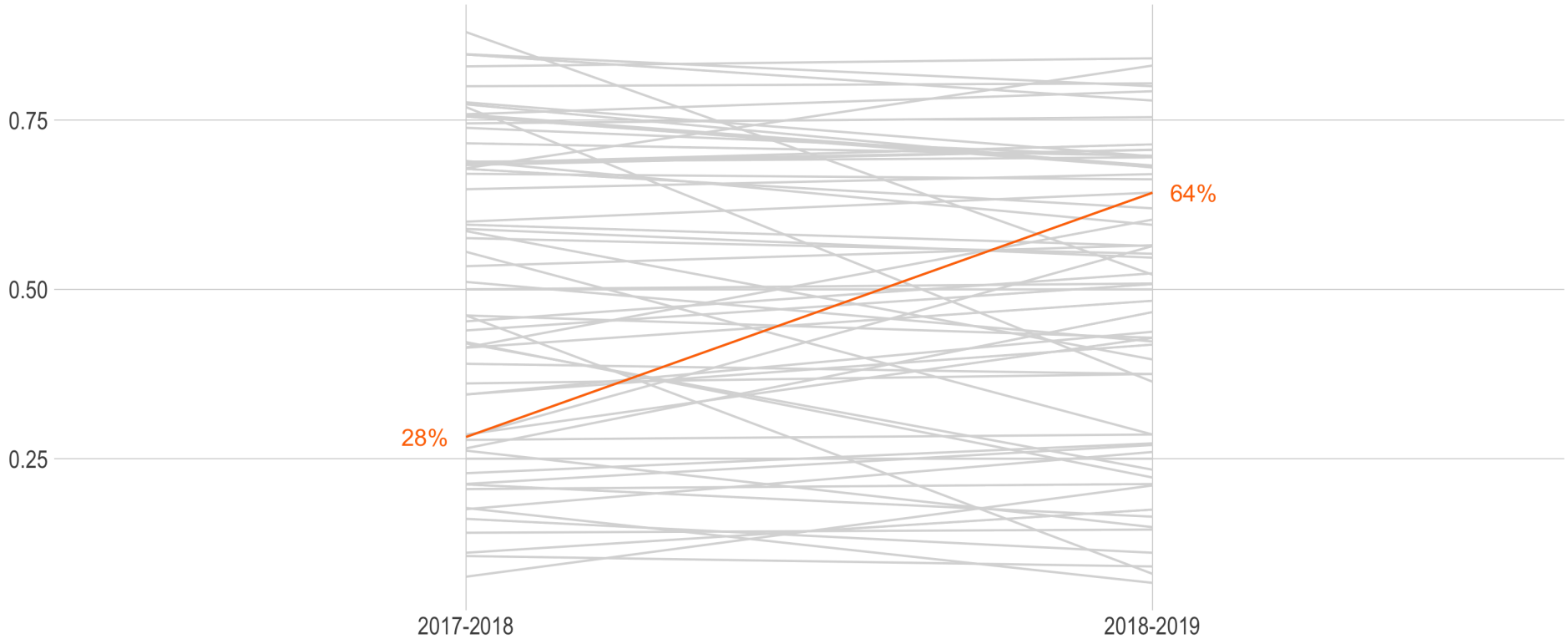
Use Direct Labeling


Most western countries are on the same coronavirus trajectory. Hong Kong and Singapore have managed to slow the spread
Cumulative number of cases, by number of days since 100th case
33% DAILY INCREASE
Italy
Iran
S. Korea: spread slowed from initial pace, but has picked up again with an outbreak in Seoul
US
Spain
France
Germany
Switzerland
Norway
Netherlands
UK
Sweden
Belgium
Japan: there is debate over whether the spread has slowed or there are simply not enough tests being done
Singapore: very strict quarantine rules and rigorous contact tracing
Hong Kong: rapid school closures and quarantining, strong community response
10,000
5,000
2,000
1,000
500
200
100
0
5
10
15
20
Number of days since 100th case
Source: FT analysis of Johns Hopkins University, CSSE
FT graphic: John Burn-Murdoch / @jburnmurdoch
© FT

Use Direct Labeling

```
rru_gray <- "#d9d9d9"

third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  geom_text(data = highlight_school,
            inherit.aes = TRUE,
            aes(label = percent_proficient_display),
            color = rru_orange,
            nudge_x = c(-0.06, 0.06)) +
  theme_ipsum(axis_title_size = 0) +
  theme(axis.title = element_blank(),
        panel.grid.minor = element_blank())
```

Use Direct Labeling



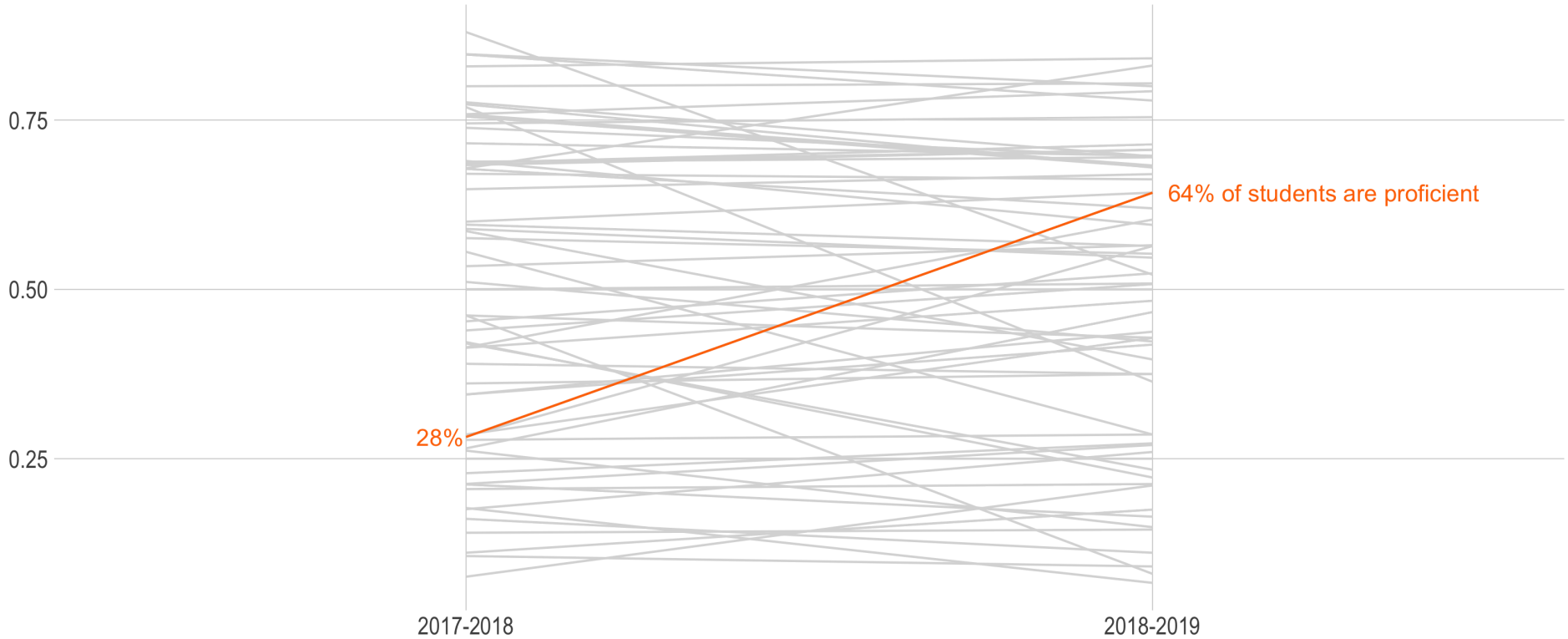
Use Direct Labeling

```
highlight_school <- third_grade_math_proficiency %>%
  filter(school == "Vestal Elementary School") %>%
  mutate(percent_proficient_display = percent(percent_proficient, accuracy = 1)) %>%
  mutate(percent_proficient_display = case_when(
    year == "2018-2019" ~ str_glue("{percent_proficient_display} of students are proficient"),
    TRUE ~ percent_proficient_display
  ))
```

Use Direct Labeling

```
third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  geom_text(data = highlight_school,
            inherit.aes = TRUE,
            aes(label = percent_proficient_display),
            color = rru_orange,
            hjust = c(1.05, -0.05)) +
  theme_ipsum(axis_title_size = 0) +
  theme(axis.title = element_blank(),
        panel.grid.minor = element_blank())
```

Use Direct Labeling

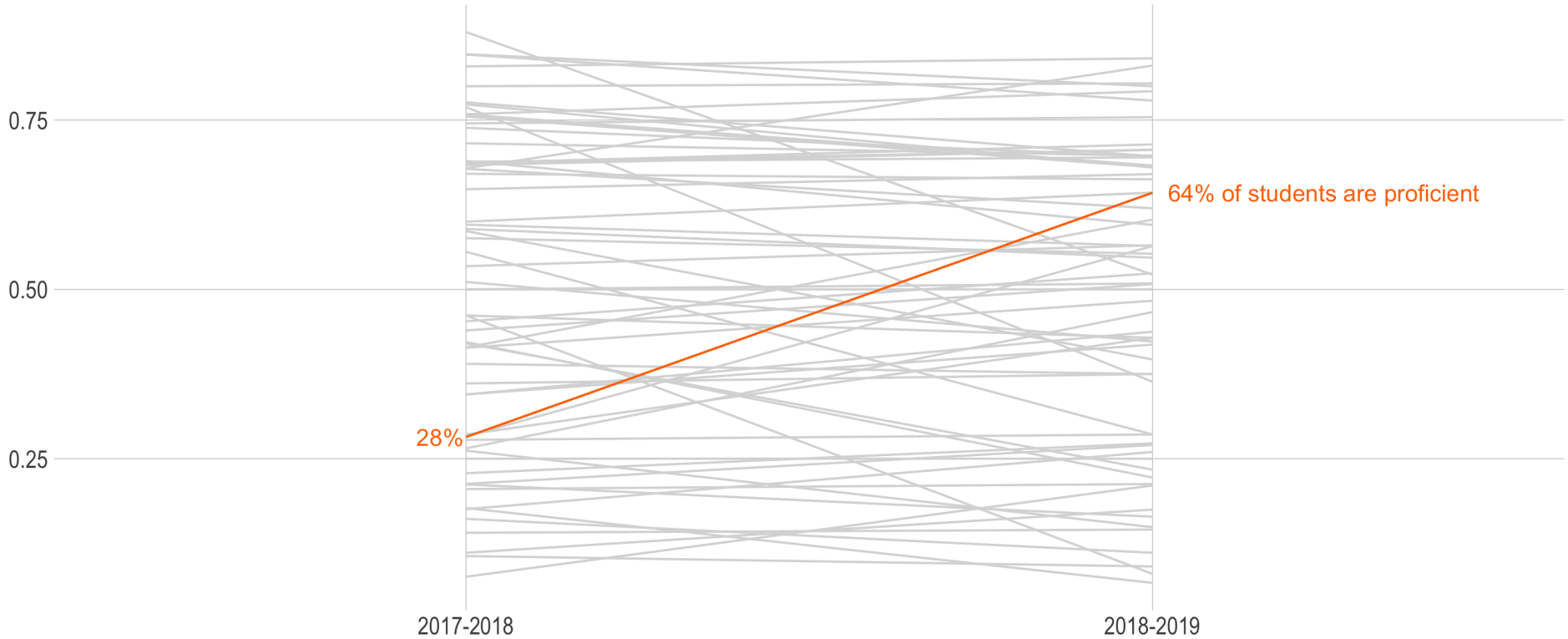


Your Turn

Add text to display the percentage of Hispanic/Latino students in Douglas ESD in 2017-2018 and 2018-2019

Use Axis Text Wisely

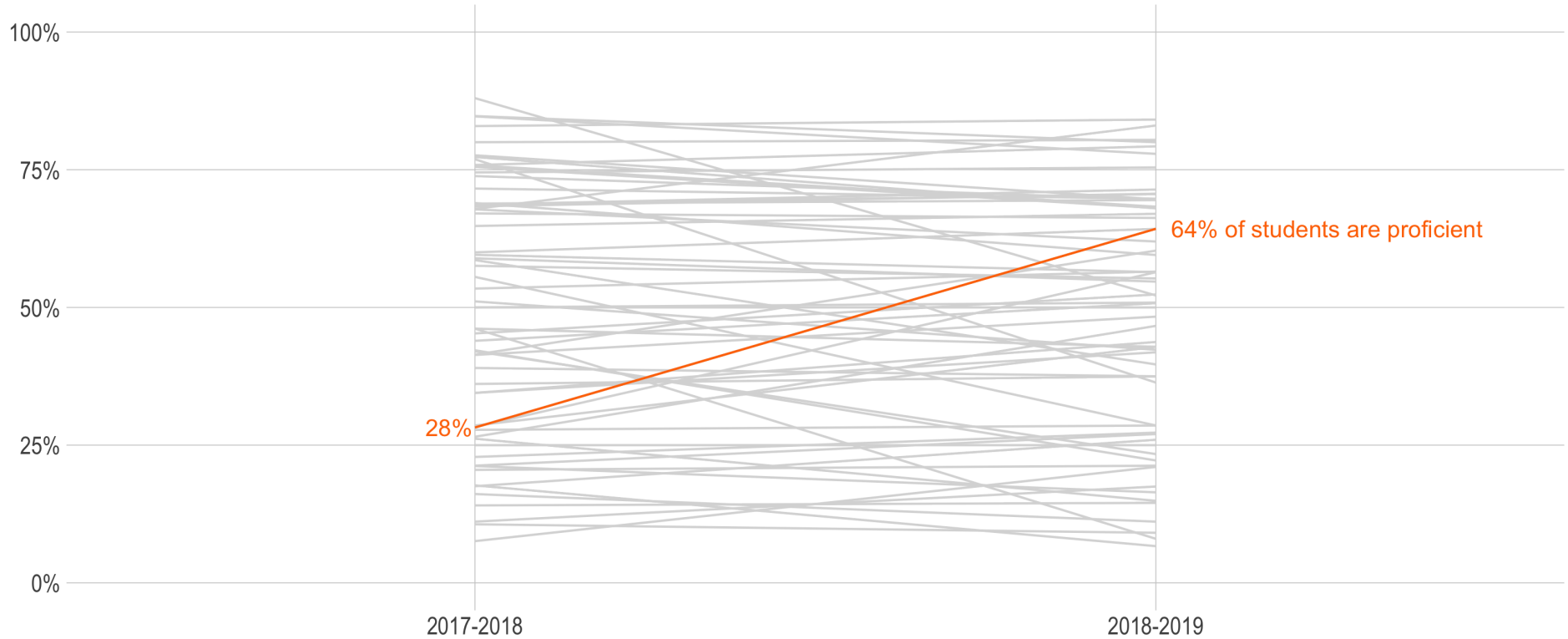
Use Axis Text Wisely



Use Axis Text Wisely

```
third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  geom_text(data = highlight_school,
            inherit.aes = TRUE,
            aes(label = percent_proficient_display),
            color = rru_orange,
            hjust = c(1.05, -0.05)) +
  scale_y_continuous(labels = percent_format(),
                     limits = c(0, 1)) +
  theme_ipsum(axis_title_size = 0,
              grid = "XY")
```

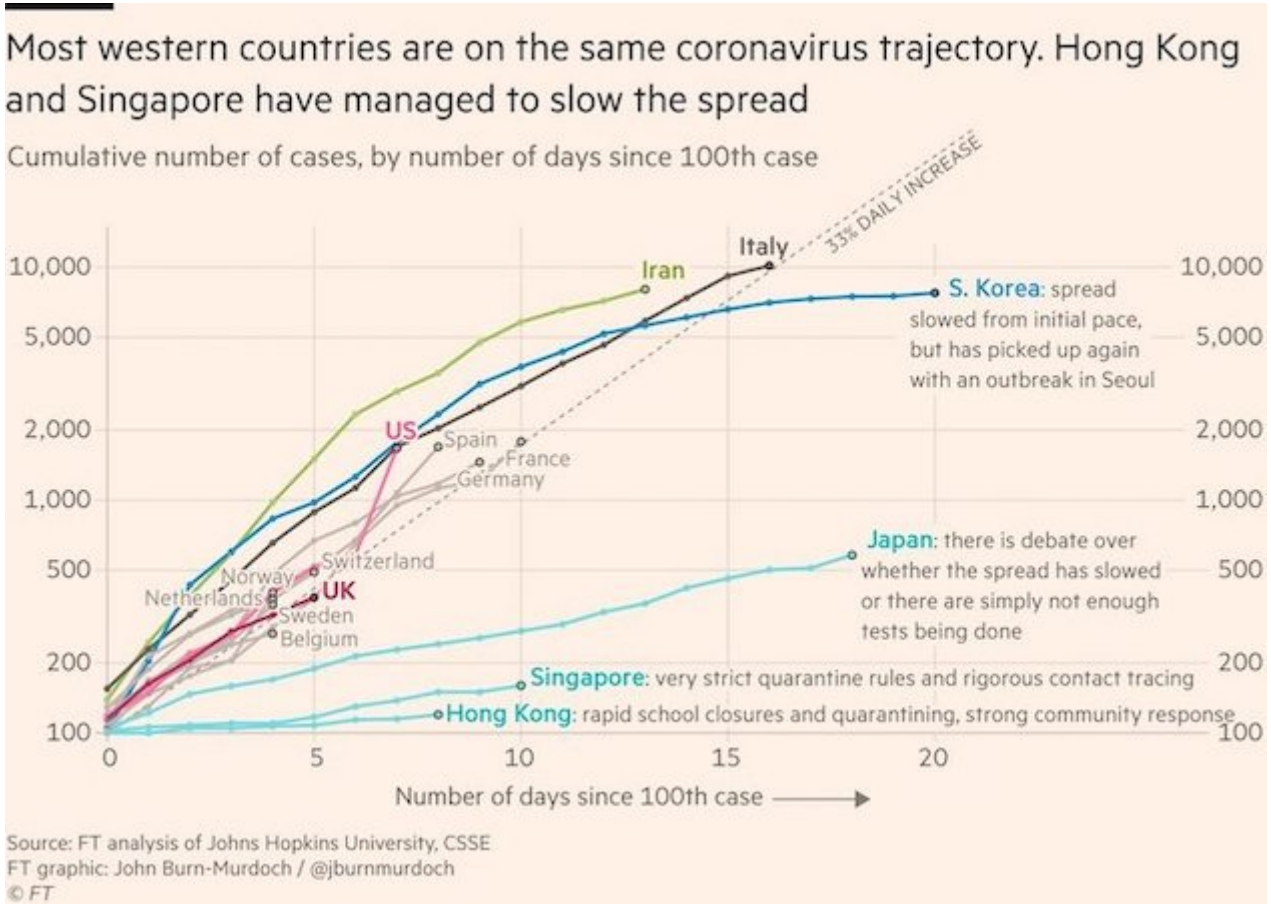
Use Axis Text Wisely



Your Turn

Make your y axis labels show up as nicely formatted percents using the `percent_format()` function.

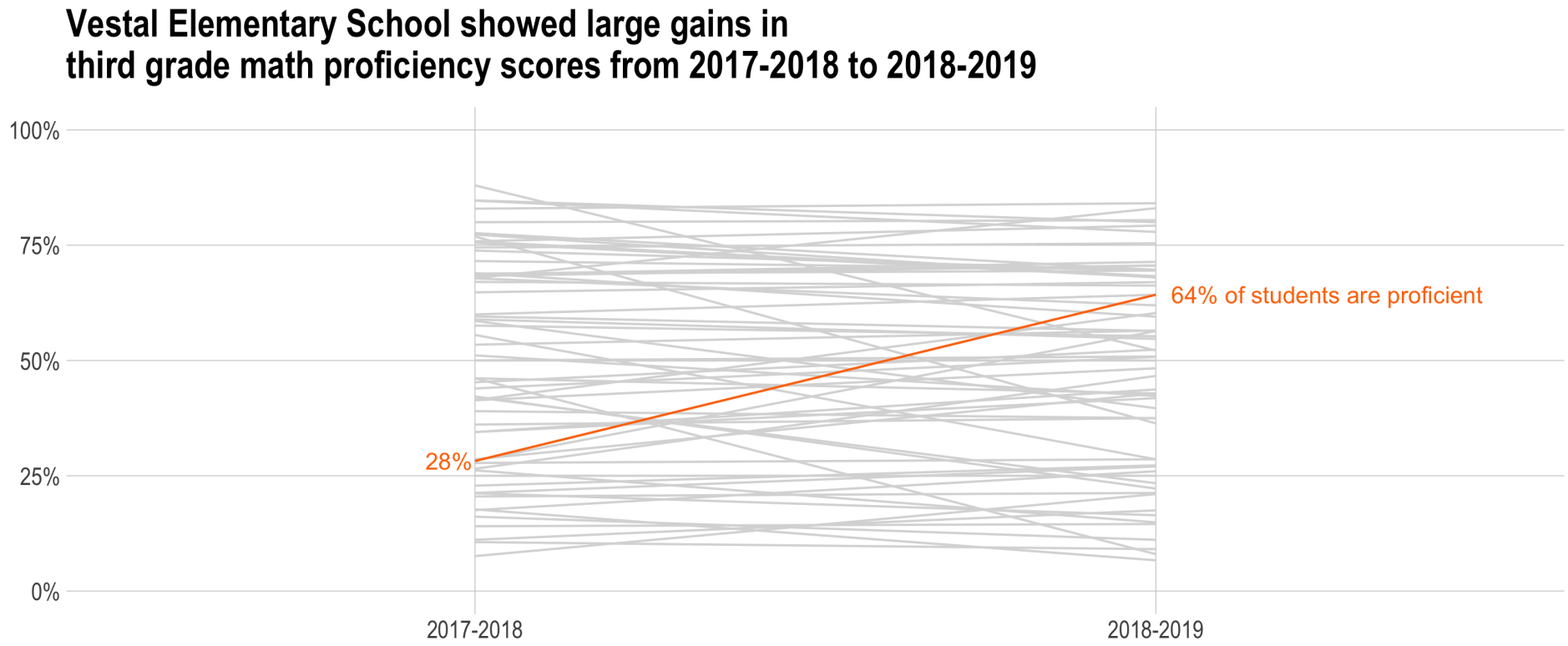
Use Titles to Highlight Findings


Most western countries are on the same coronavirus trajectory. Hong Kong and Singapore have managed to slow the spread
Cumulative number of cases, by number of days since 100th case
33% DAILY INCREASE
Italy
Iran
S. Korea: spread slowed from initial pace, but has picked up again with an outbreak in Seoul
US
Spain
France
Germany
Switzerland
Norway
Netherlands
UK
Sweden
Belgium
Japan: there is debate over whether the spread has slowed or there are simply not enough tests being done
Singapore: very strict quarantine rules and rigorous contact tracing
Hong Kong: rapid school closures and quarantining, strong community response
10,000
5,000
2,000
1,000
500
200
100
0
5
10
15
20
Number of days since 100th case
Source: FT analysis of Johns Hopkins University, CSSE
FT graphic: John Burn-Murdoch / @jburnmurdoch
© FT

Use Titles to Highlight Findings

```
third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  geom_text(data = highlight_school,
            inherit.aes = TRUE,
            aes(label = percent_proficient_display),
            color = rru_orange,
            hjust = c(1.05, -0.05)) +
  scale_y_continuous(labels = percent_format(accuracy = 1)) +
  labs(title = "Vestal Elementary School showed large gains in\nthird grade math proficiency scores from 2017-2018 to 2018-2019")
  scale_y_continuous(labels = percent_format(),
                     limits = c(0, 1)) +
  theme_ipsum(axis_title_size = 0,
              grid = "XY")
```

Use Titles to Highlight Findings



Your Turn

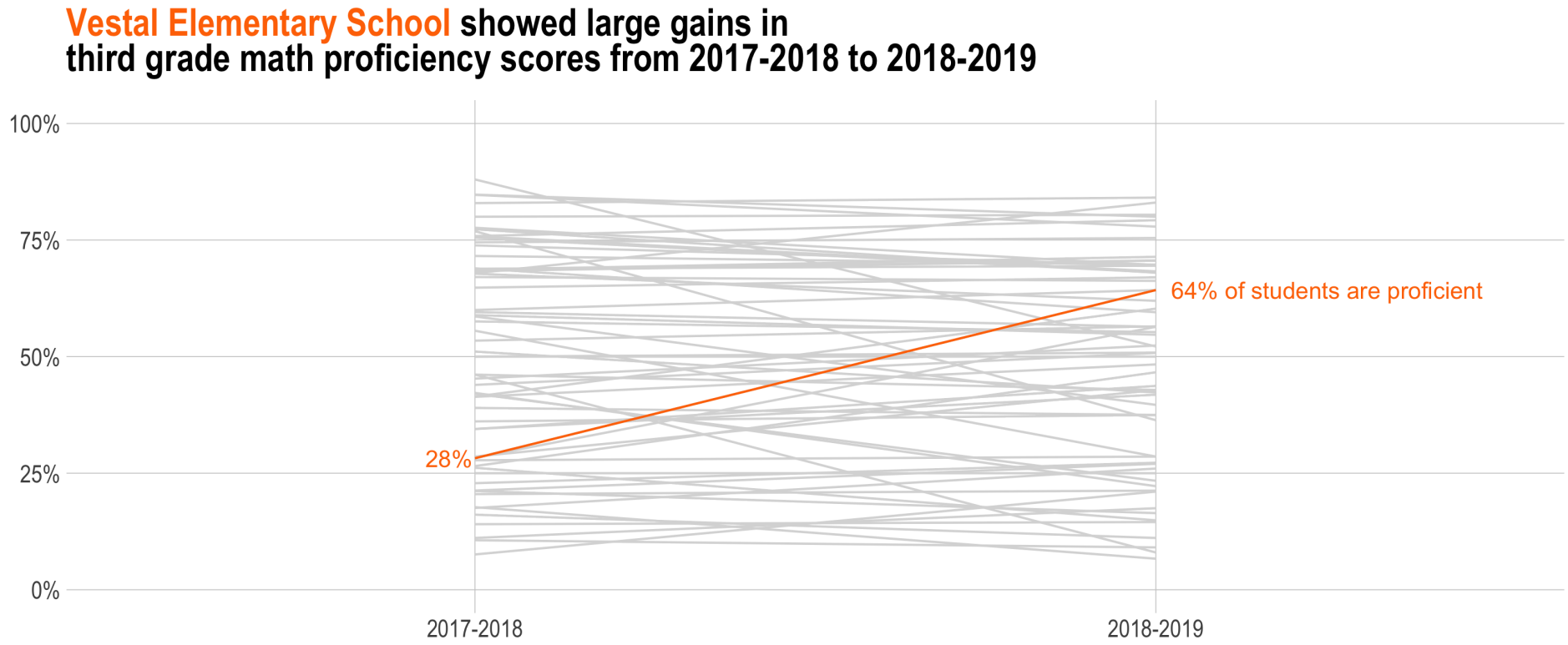
Add a title to highlight your main finding


Use Color in Titles to
Highlight Findings

Use Color in Titles to Highlight Findings

```
library(ggtext)

third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  geom_text(data = highlight_school,
            inherit.aes = TRUE,
            aes(label = percent_proficient_display),
            color = rru_orange,
            hjust = c(1.05, -0.05)) +
  labs(title = "<span style = 'color: #FF7400;'>Vestal Elementary School</span> showed large gains in<br>third grade math profici
  scale_y_continuous(labels = percent_format(),
                     limits = c(0, 1)) +
  theme_ipsum(axis_title_size = 0,
              grid = "XY") +
  theme(plot.title = element_markdown())
```

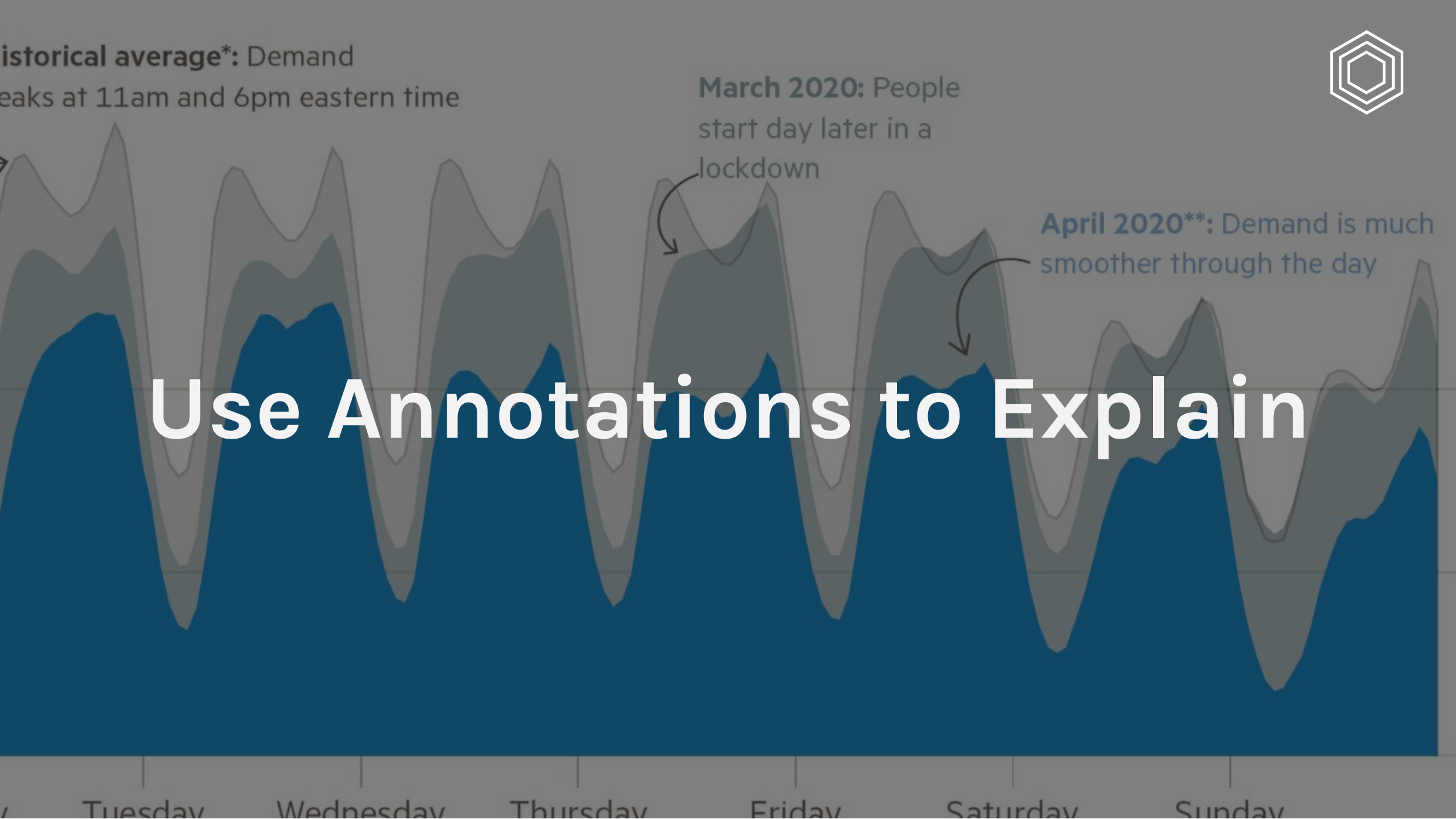
Use Color in Titles to Highlight Findings



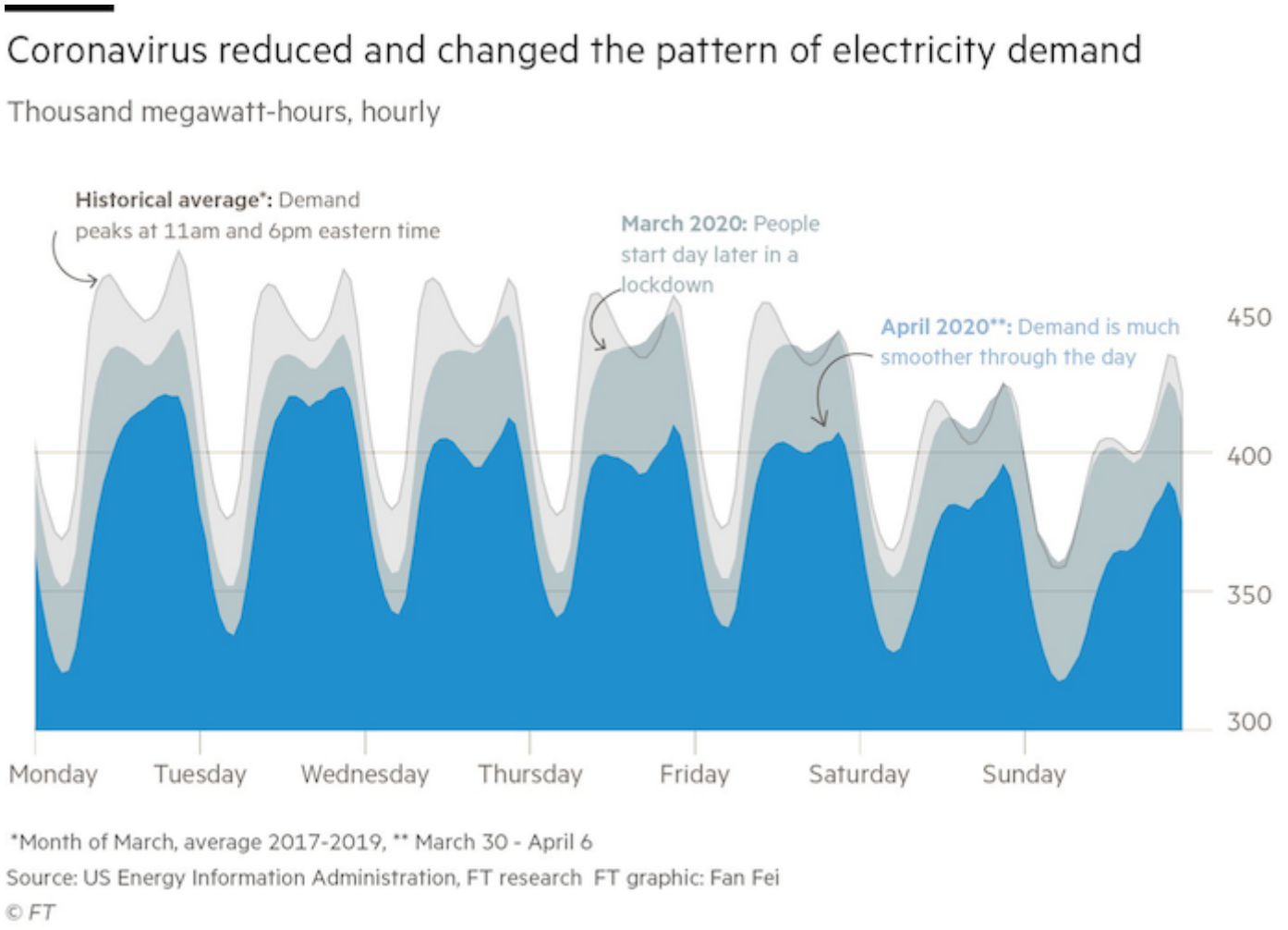
Your Turn

Use color in your title to highlight your main finding. You'll need to:

1. Add HTML in the `labs()` function to add the title
2. Change the `plot.title` argument in the `theme()` function so that it interprets the HTML correctly



Use Annotations to Explain

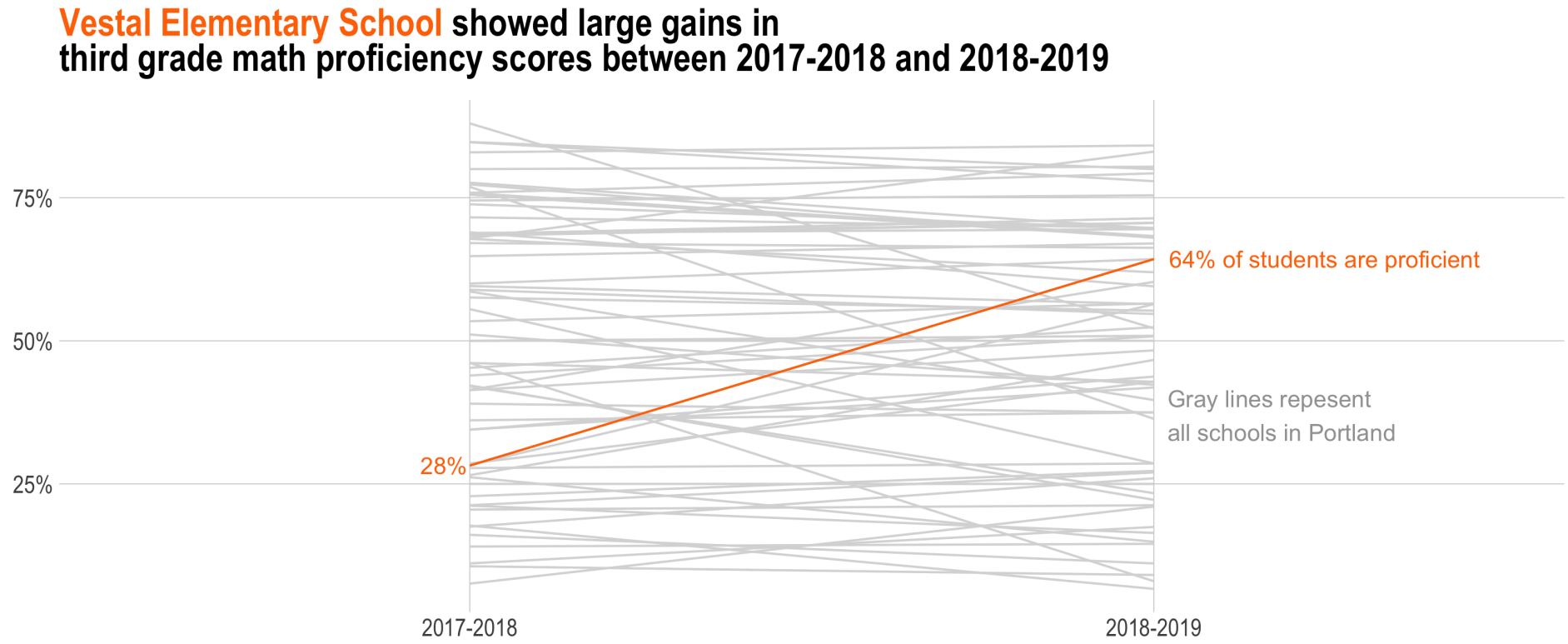


Source: [Financial Times]

Use Annotations to Explain

```
third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  geom_text(data = highlight_school,
            inherit.aes = TRUE,
            aes(label = percent_proficient_display),
            color = rru_orange,
            hjust = c(1.05, -0.05)) +
  scale_y_continuous(labels = percent_format(accuracy = 1)) +
  theme_ipsum(axis_title_size = 0,
              grid = "XY") +
  labs(title = "<span style = 'color: #FF7400'>Vestal Elementary School</span> showed large gains in<br>third grade math proficie
  theme(plot.title = element_markdown()) +
  annotate("text",
           x = 2.02,
           y = .37,
           label = "Gray lines repesent\nall schools in Portland",
           color = "#A0A0A0",
           hjust = 0)
```

Use Annotations to Explain

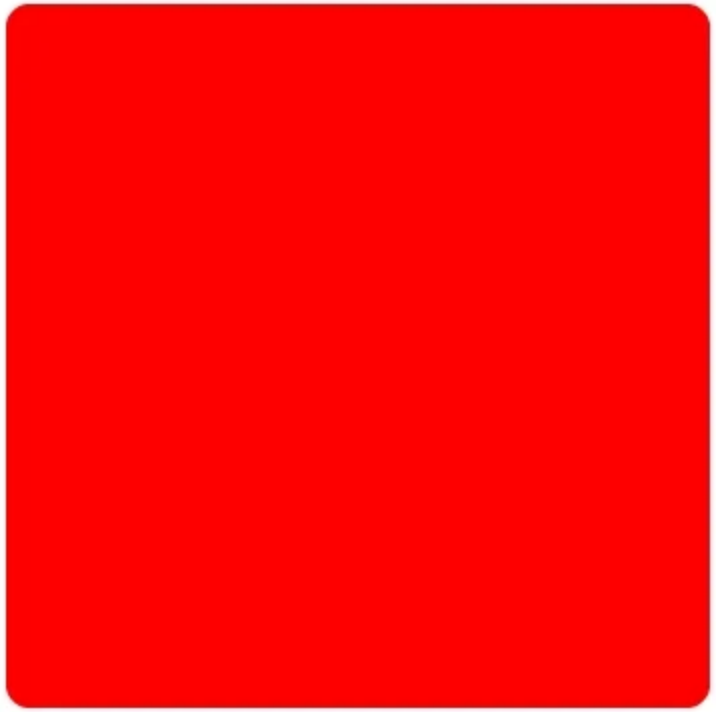


Your Turn

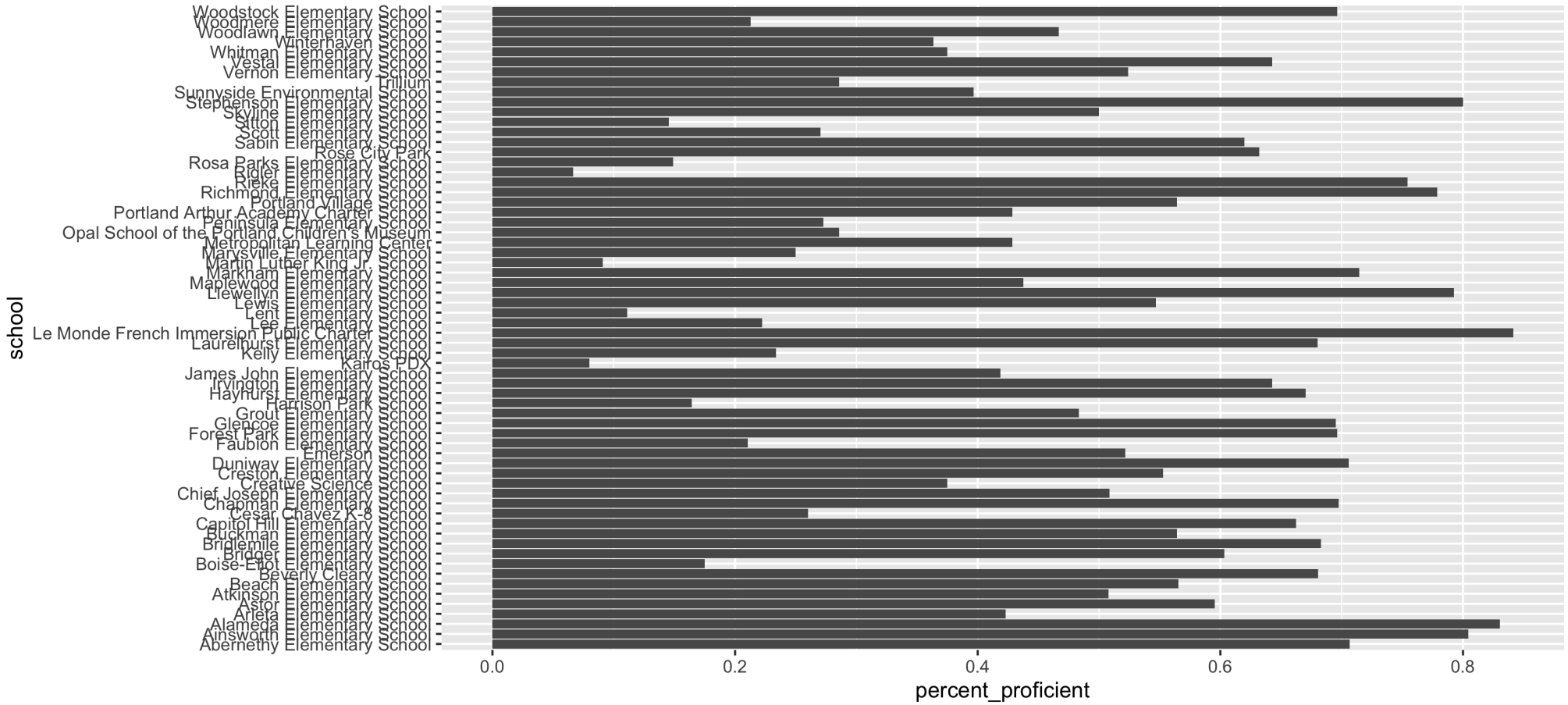
Add an annotation somewhere on your chart to help the reader understand it better


Make it Sparkle

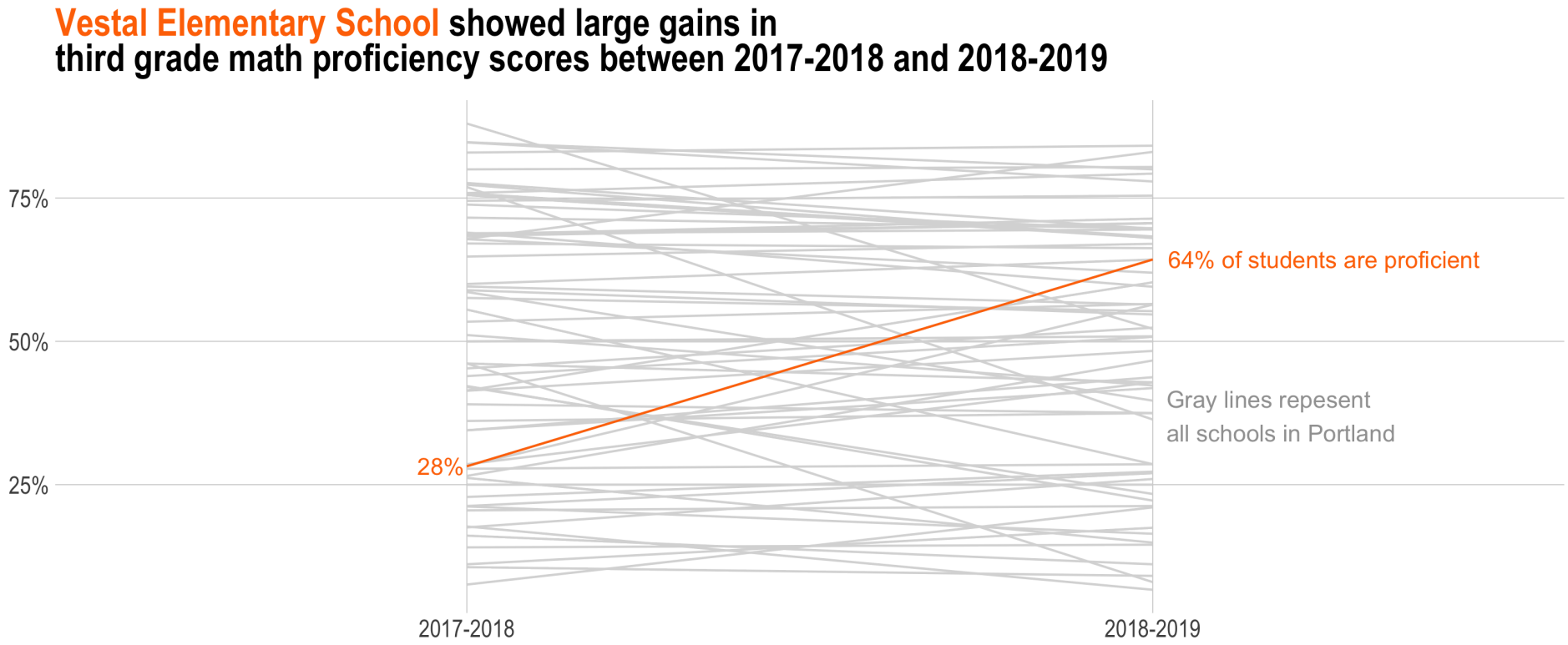
Don't Use Defaults



Don't Use Defaults



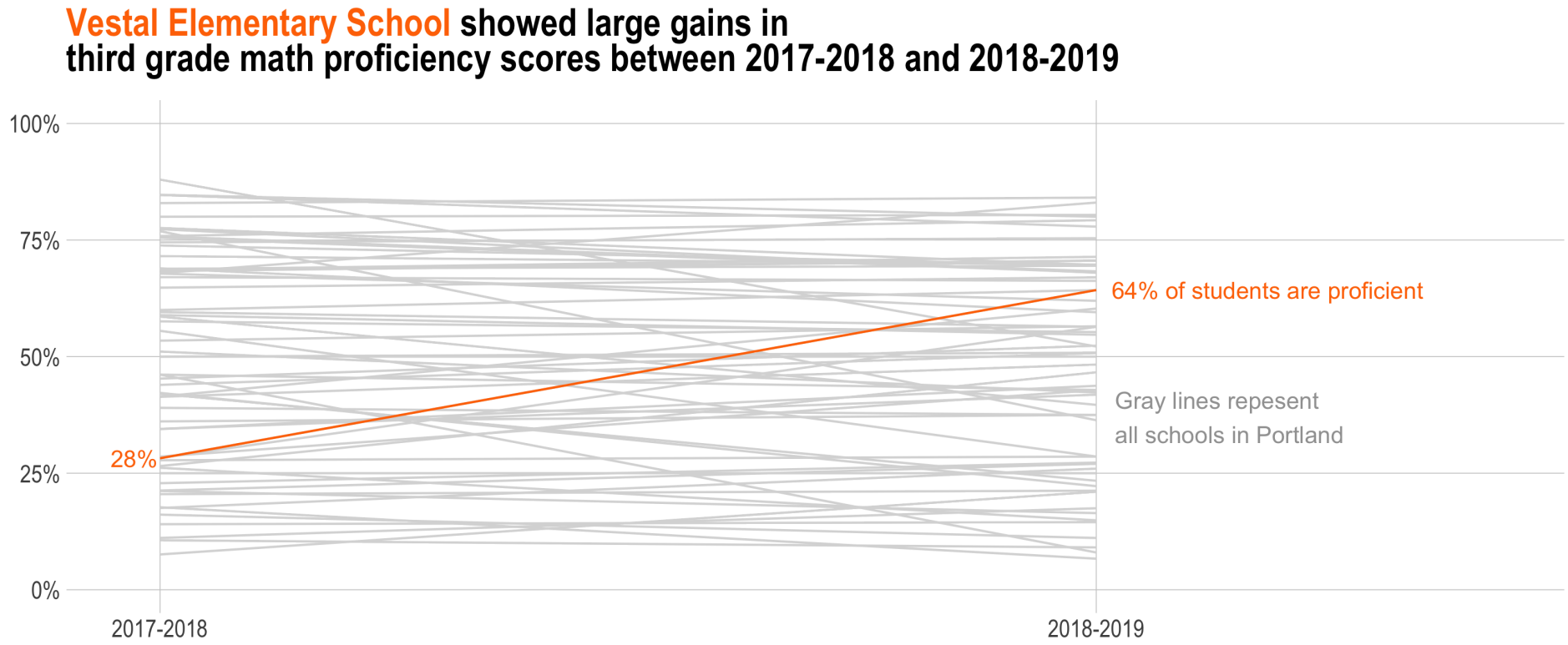
Tweak Spacing Around Your Plot



Tweak Spacing Around Your Plot

```
third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  geom_text(data = highlight_school,
            inherit.aes = TRUE,
            aes(label = percent_proficient_display),
            color = rru_orange,
            hjust = c(1.05, -0.05)) +
  scale_y_continuous(label = percent_format(),
                     limits = c(0, 1)) +
  theme_ipsum(axis_title_size = 0,
              grid = "XY") +
  labs(title = "<span style = 'color: #FF7400'>Vestal Elementary School</span> showed large gains in<br>third grade math proficie
  theme(plot.title = element_markdown()) +
  annotate("text",
           x = 2.02,
           y = .37,
           label = "Gray lines repesent\nall schools in Portland",
           color = "#A0A0A0",
           hjust = 0) +
  scale_x_discrete(expand = expansion(add = c(0.1, 0.5)))
```

Tweak Spacing Around Your Plot



Your Turn

Use the expand argument within the `scale_x_discrete()` function to remove any gaps in your plot

Customize Your Theme

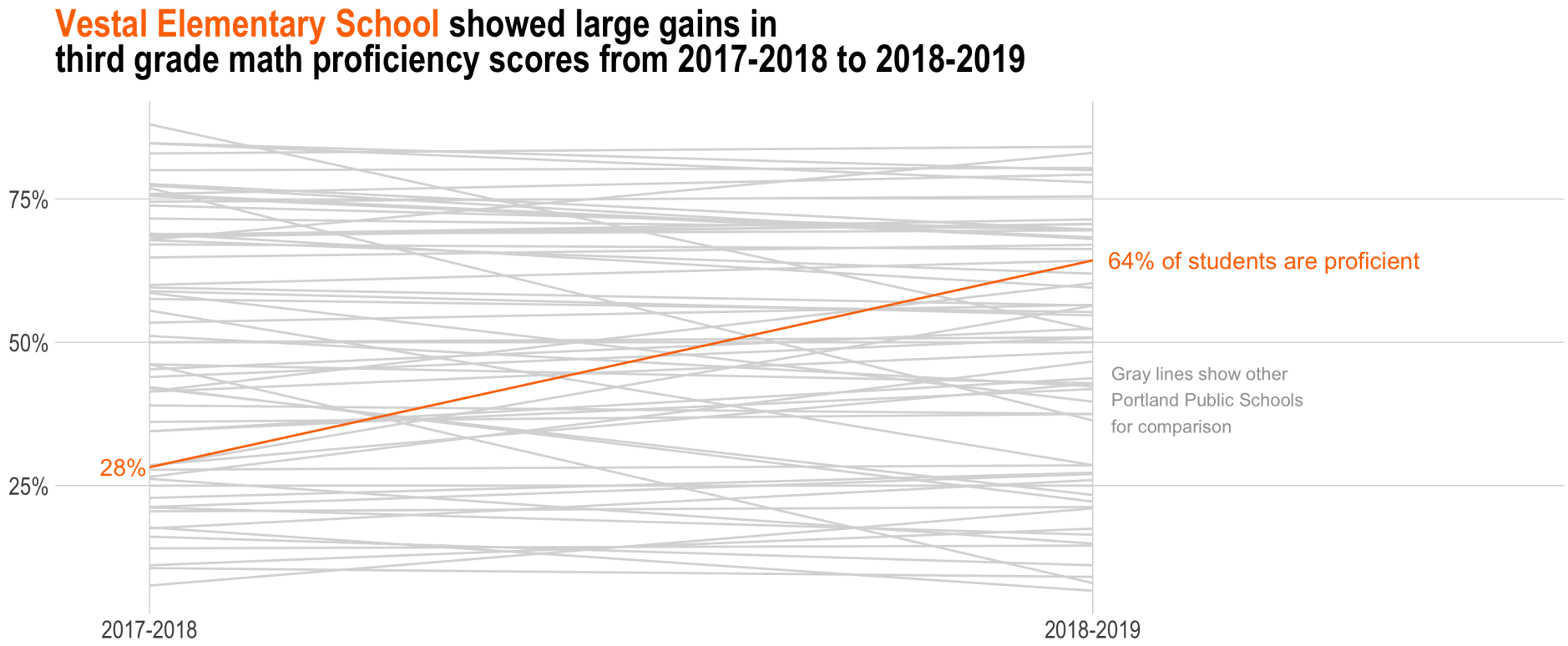
Make Your Own Custom Theme

```
theme_dk <- function() {
  theme_ipsum(axis_title_size = 0,
              grid = "XY") +
    theme(plot.title = element_markdown())
}
```

Make Your Own Custom Theme

```
third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  geom_text(data = highlight_school,
            inherit.aes = TRUE,
            aes(label = percent_proficient_display),
            color = rru_orange,
            hjust = c(1.05, -0.05)) +
  scale_y_continuous(labels = percent_format(accuracy = 1)) +
  labs(title = "<span style = 'color: #FF7400;'>Vestal Elementary School</span> showed large gains in<br>third grade math profici
  annotate("text",
           x = 2.02,
           y = 0.4,
           label = "Gray lines show other\nPortland Public Schools\nfor comparison",
           hjust = 0,
           size = 3,
           color = "#A0A0A0") +
  scale_x_discrete(expand = expansion(add = c(0.1, 0.5))) +
  theme_dk()
```

Make Your Own Custom Theme



Your Turn

1. Make your own theme by combining elements of the `theme()` function and/or themes from other packages (e.g. `theme_ipsum()`)

2. Add this theme to your plot, removing any code that is now redundant

Customize Your Fonts

Import Custom Fonts

```
library(extrafont)

font_import(pattern = "Karla")
```

Load Fonts

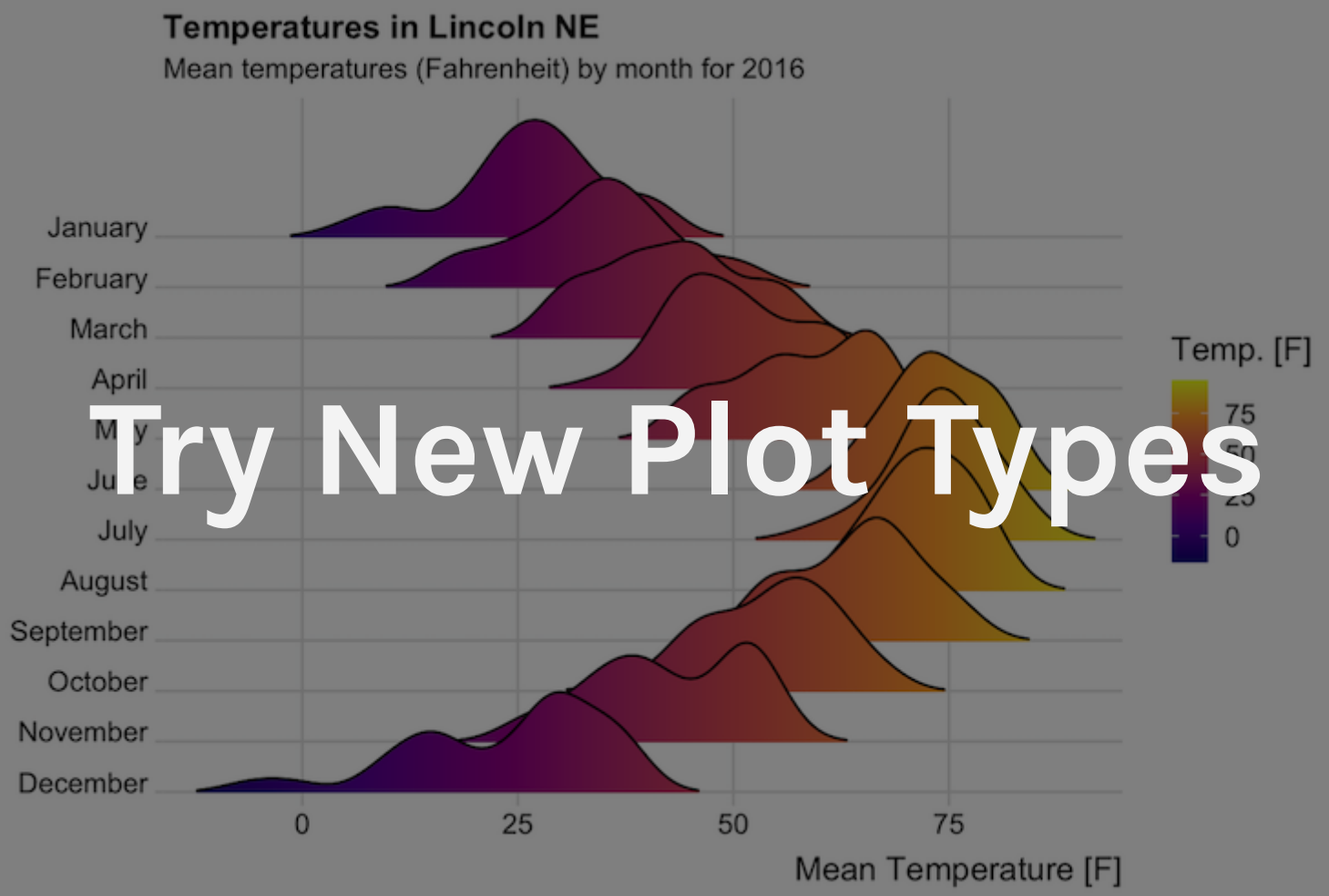
```
loadfonts()
```

```
fonts()
```

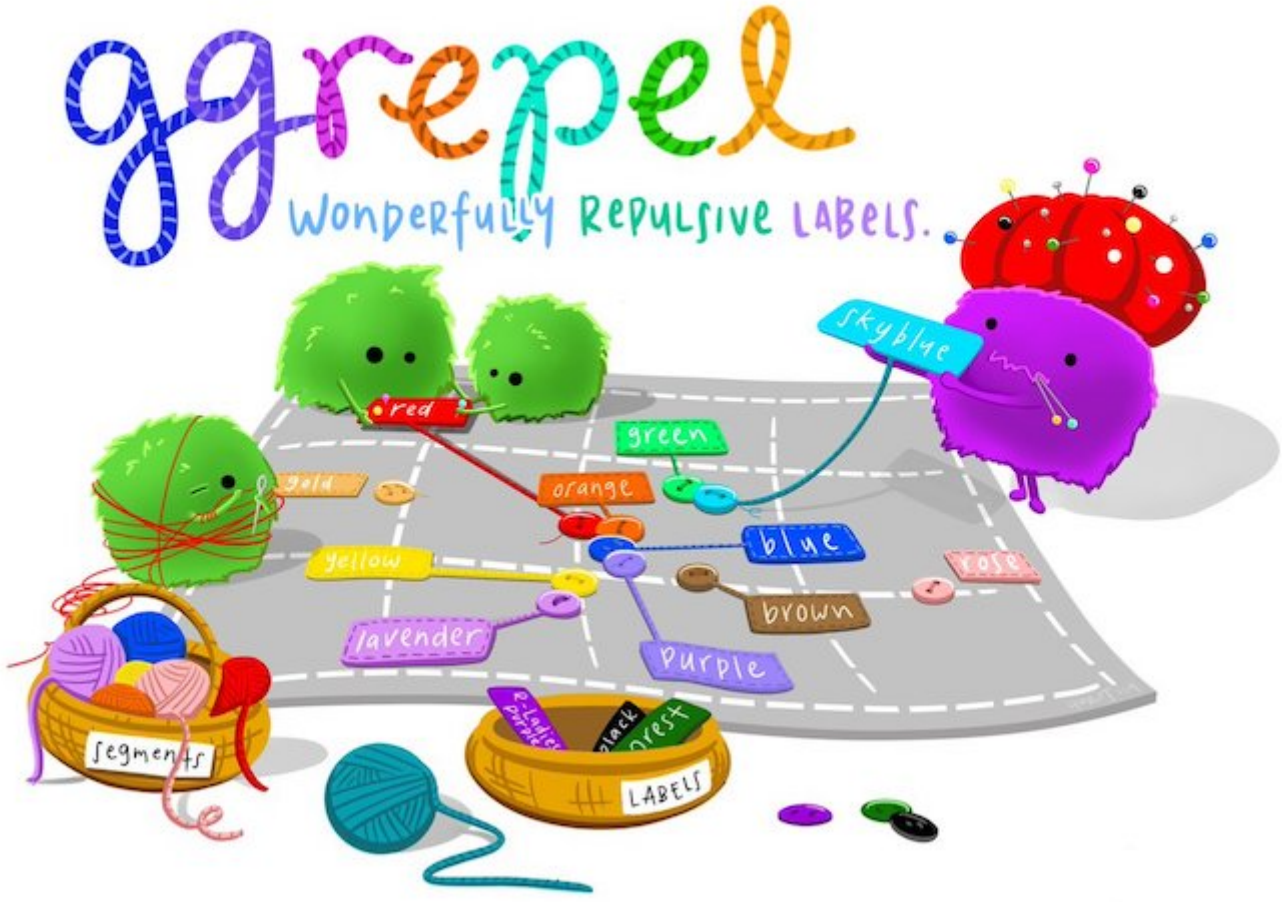
```
theme_dk <- function() {
  theme_ipsum(axis_title_size = 0,
              base_family = "Karla",
              grid = "XY") +
    theme(plot.title = element_markdown())
}

third_grade_math_proficiency %>%
  filter(district == "Portland SD 1J") %>%
  ggplot(aes(x = year, y = percent_proficient,
             group = school)) +
  geom_line(color = rru_gray) +
  geom_line(data = highlight_school,
            inherit.aes = TRUE,
            color = rru_orange) +
  geom_text(data = highlight_school,
            inherit.aes = TRUE,
            aes(label = percent_proficient_display),
            color = rru_orange,
            family = "Karla",
            hjust = c(1.05, -0.05)) +
  scale_y_continuous(label = percent_format(),
                     limits = c(0, 1)) +
  labs(title = "<span style = 'color: #FF7400'>Vestal Elementary School</span> showed large gains in<br>third grade math proficie
  annotate("text",
           x = 2.02,
           y = .37,
           label = "Gray lines repesent\nall schools in Portland",
           color = "#A0A0A0",
           family = "Karla",
           hjust = 0) +
  scale_x_discrete(expand = expansion(add = c(0.1, 0.5))) +
  theme_dk()
```

Your Turn

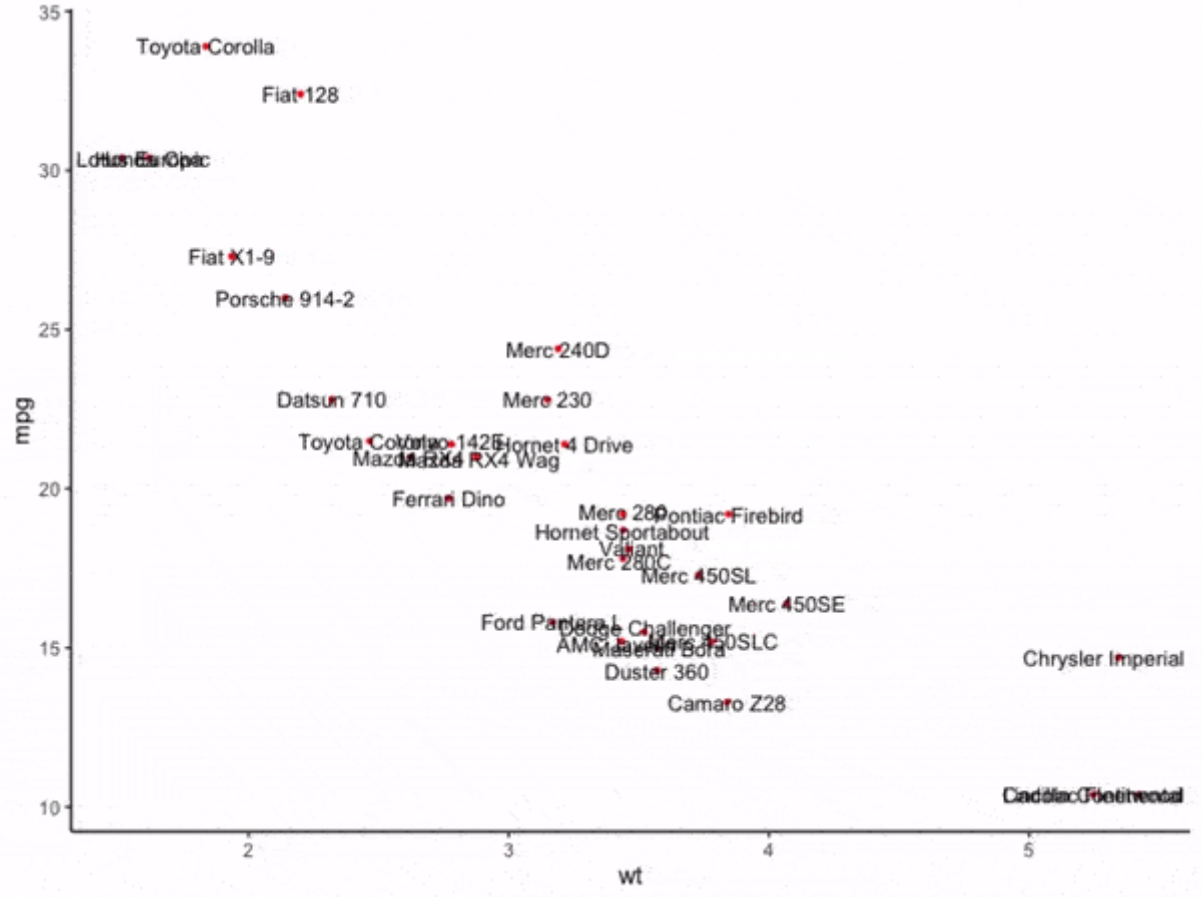
1. Install the `extrafonts` package
2. Run the `font_import()` function to make all fonts on your computer available in R (it will take a few minutes)
3. Run the `loadfonts()` function to make sure all fonts are available to R
4. Change all text to use a custom font (you'll have to do this in a few different places)



Try New Plot Types

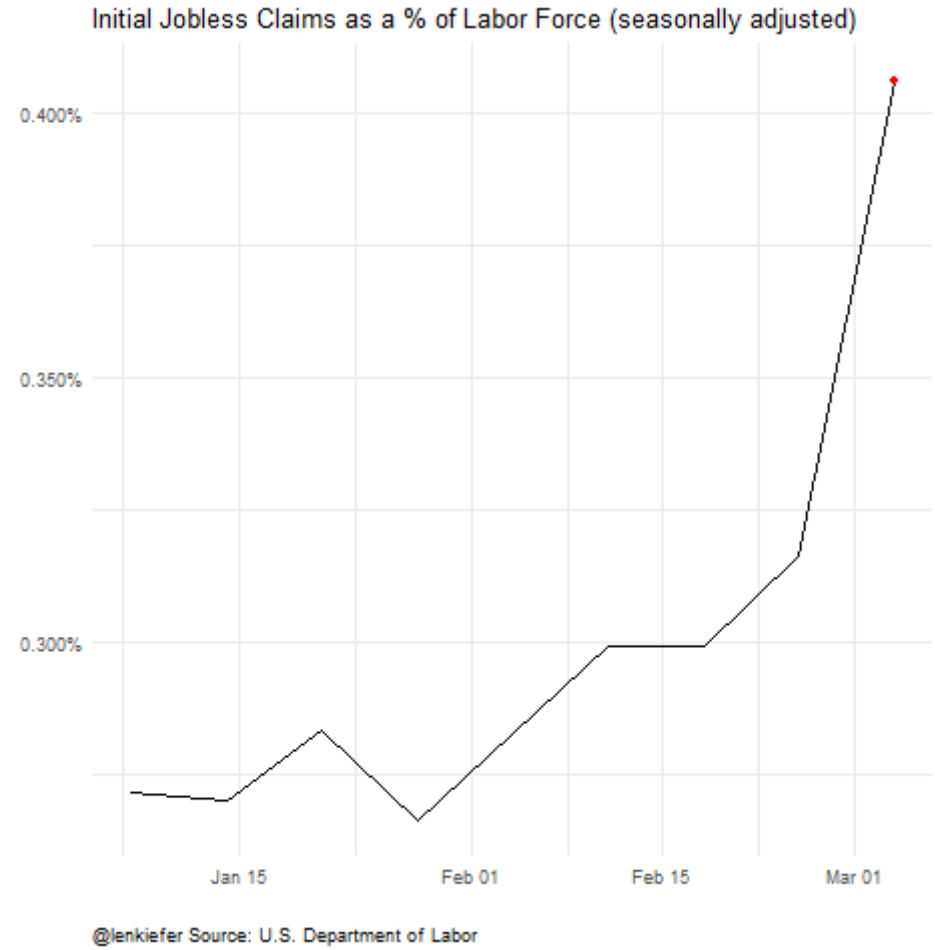

ggrepel
Wonderfully Repulsive LABELS.
red
green
gold
orange
skyblue
blue
yellow
rose
brown
lavender
purple
segments
LABELS

ggrepel

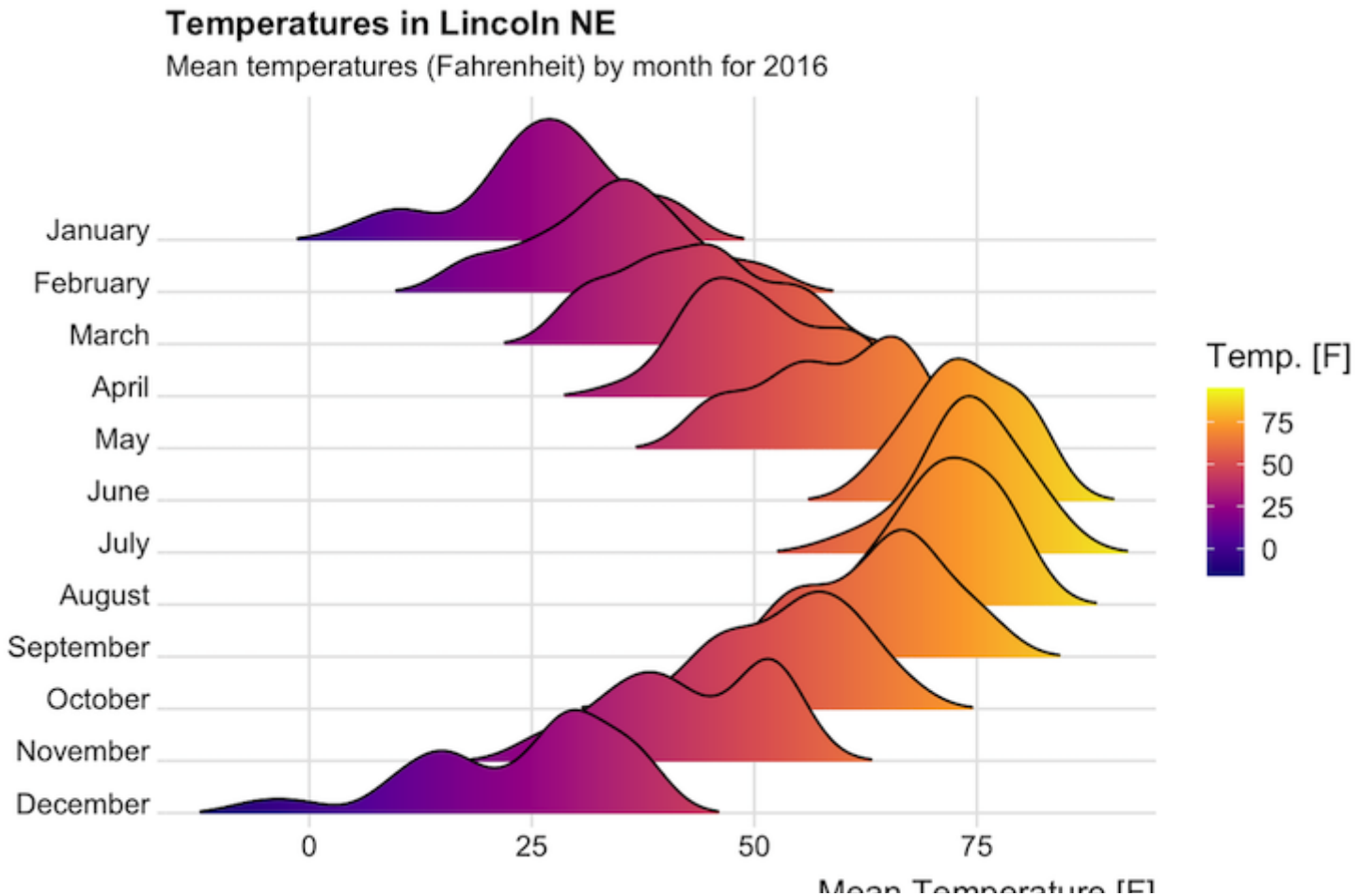

35
30
25
20
15
10
mpg
2
3
4
5
wt
Toyota Corolla
Fiat 128
Fiat X1-9
Porsche 914-2
Merc 240D
Datsun 710
Merc 230
Hornet 4 Drive
Ferrari Dino
Pontiac Firebird
Hornet Sportabout
Valiant
Merc 450SL
Merc 450SE
Dodge Challenger
Duster 360
Camaro Z28
Chrysler Imperial


gganimate:
ACTION FIGURES!
data
transition
+
enter/
Horst '18

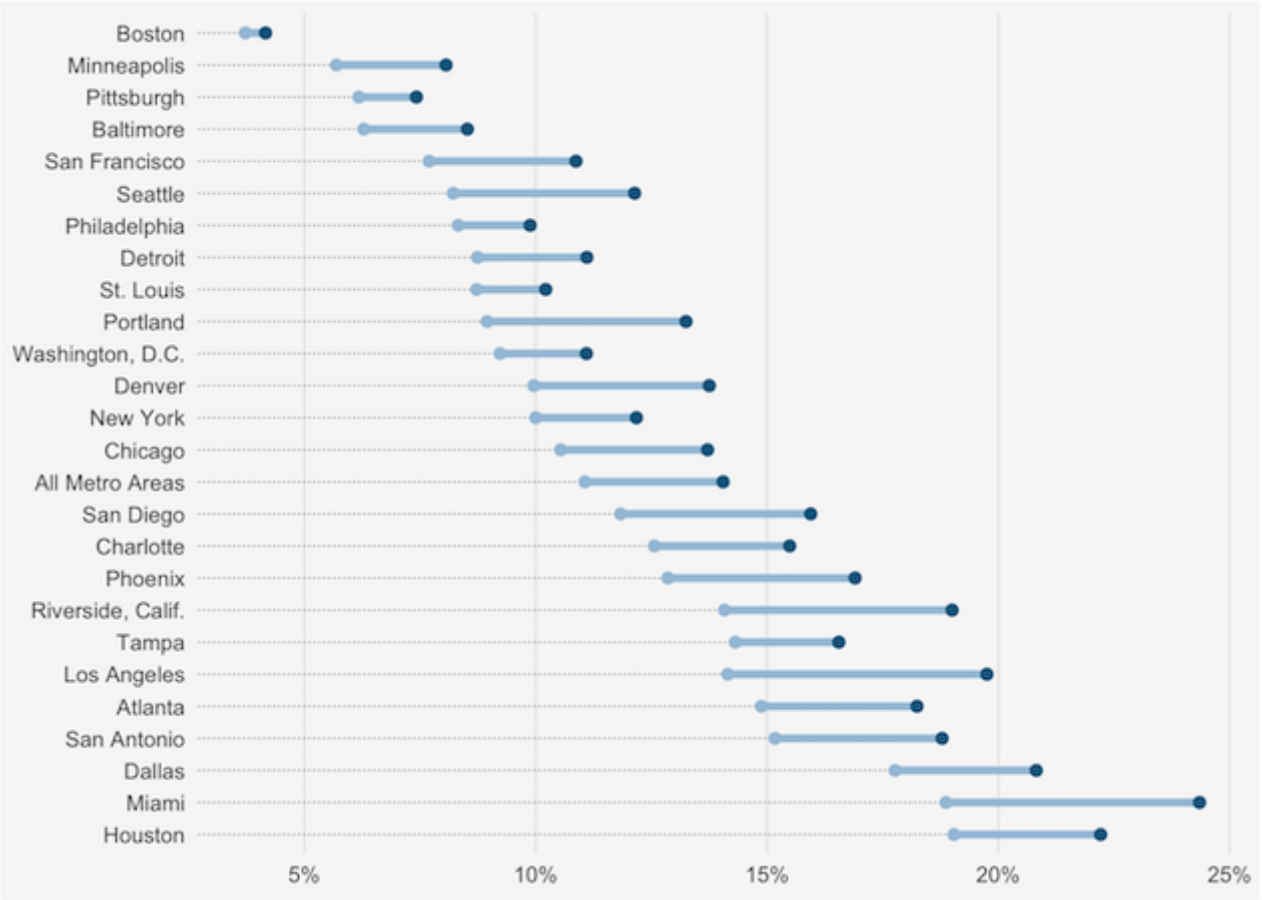
gganimate


Initial Jobless Claims as a % of Labor Force (seasonally adjusted)
0.400%
0.350%
0.300%
Jan 15
Feb 01
Feb 15
Mar 01
@lenkiefer Source: U.S. Department of Labor

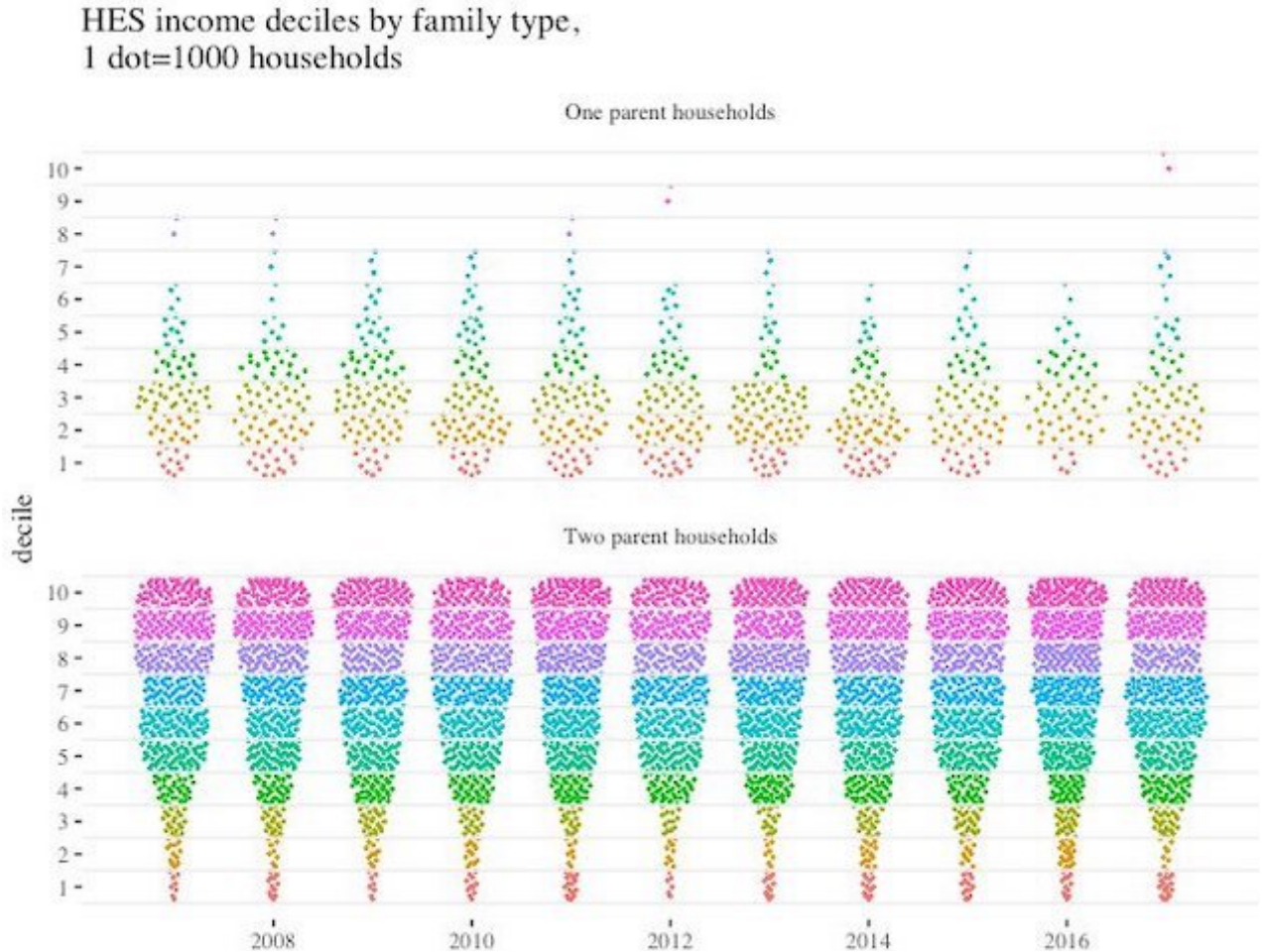
ggridges


Temperatures in Lincoln NE
Mean temperatures (Fahrenheit) by month for 2016
January
February
March
April
May
June
July
August
September
October
November
December
0
25
50
75
Temp. [F]
75
50
25
0

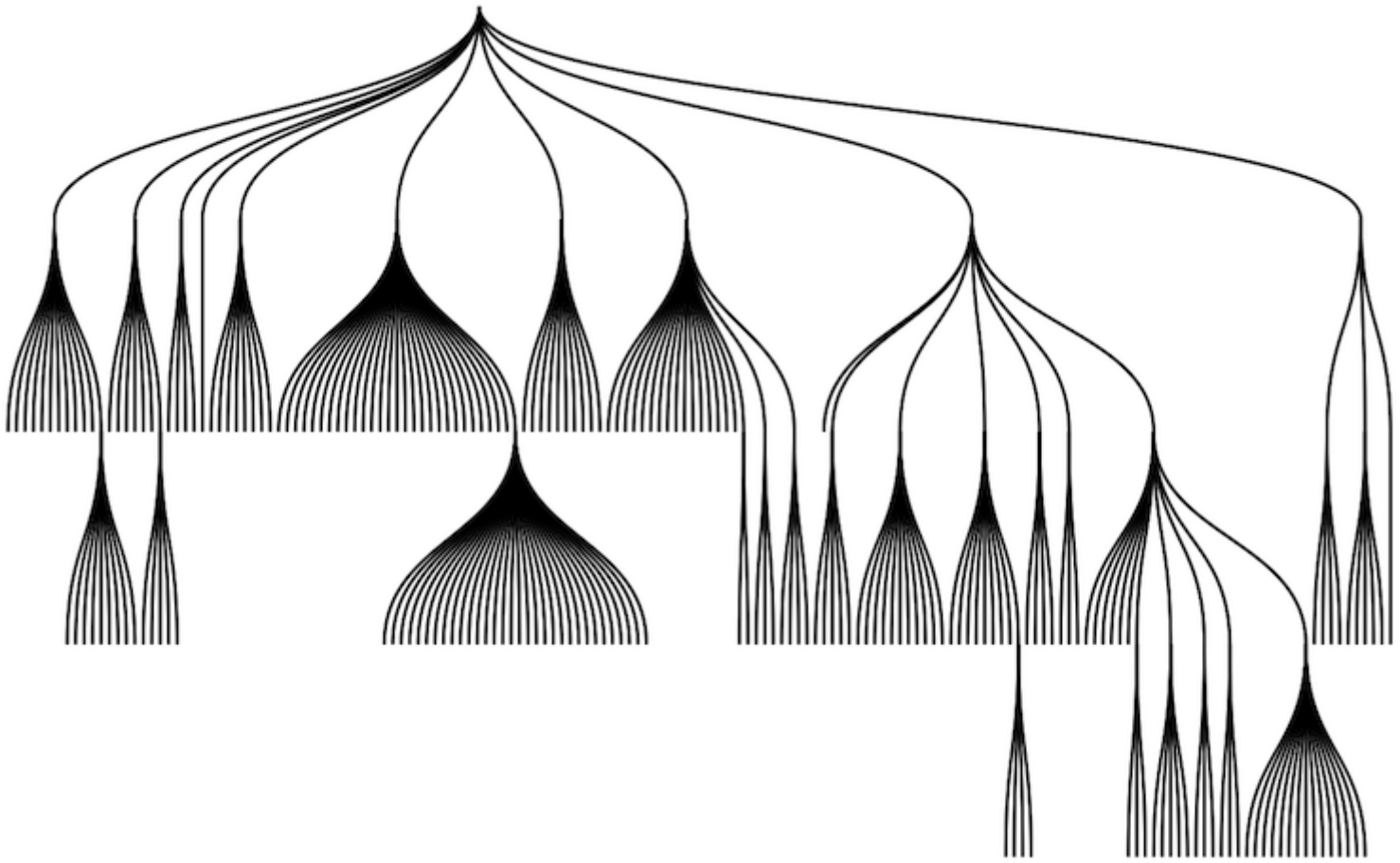
ggalt


Boston
Minneapolis
Pittsburgh
Baltimore
San Francisco
Seattle
Philadelphia
Detroit
St. Louis
Portland
Washington, D.C.
Denver
New York
Chicago
All Metro Areas
San Diego
Charlotte
Phoenix
Riverside, Calif.
Tampa
Los Angeles
Atlanta
San Antonio
Dallas
Miami
Houston
5%
10%
15%
20%
25%

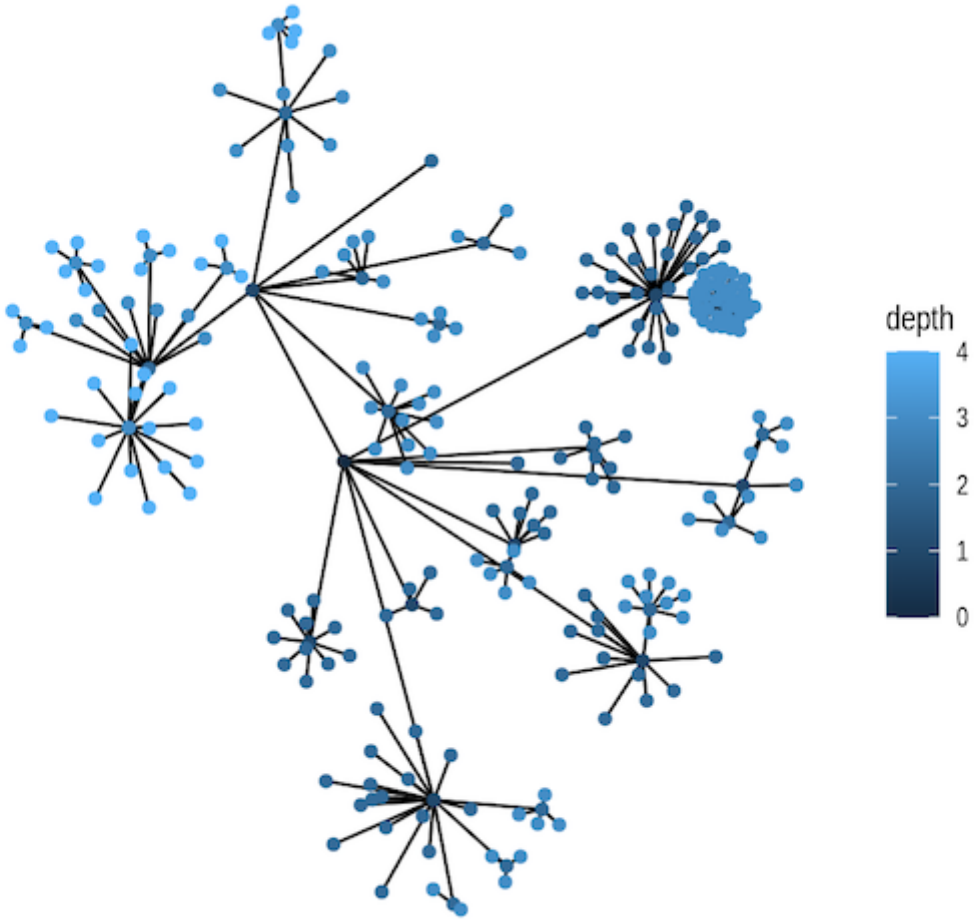
ggbeeswarm


HES income deciles by family type,
1 dot=1000 households
One parent households
Two parent households
decile
10
9
8
7
6
5
4
3
2
1
2008
2010
2012
2014
2016

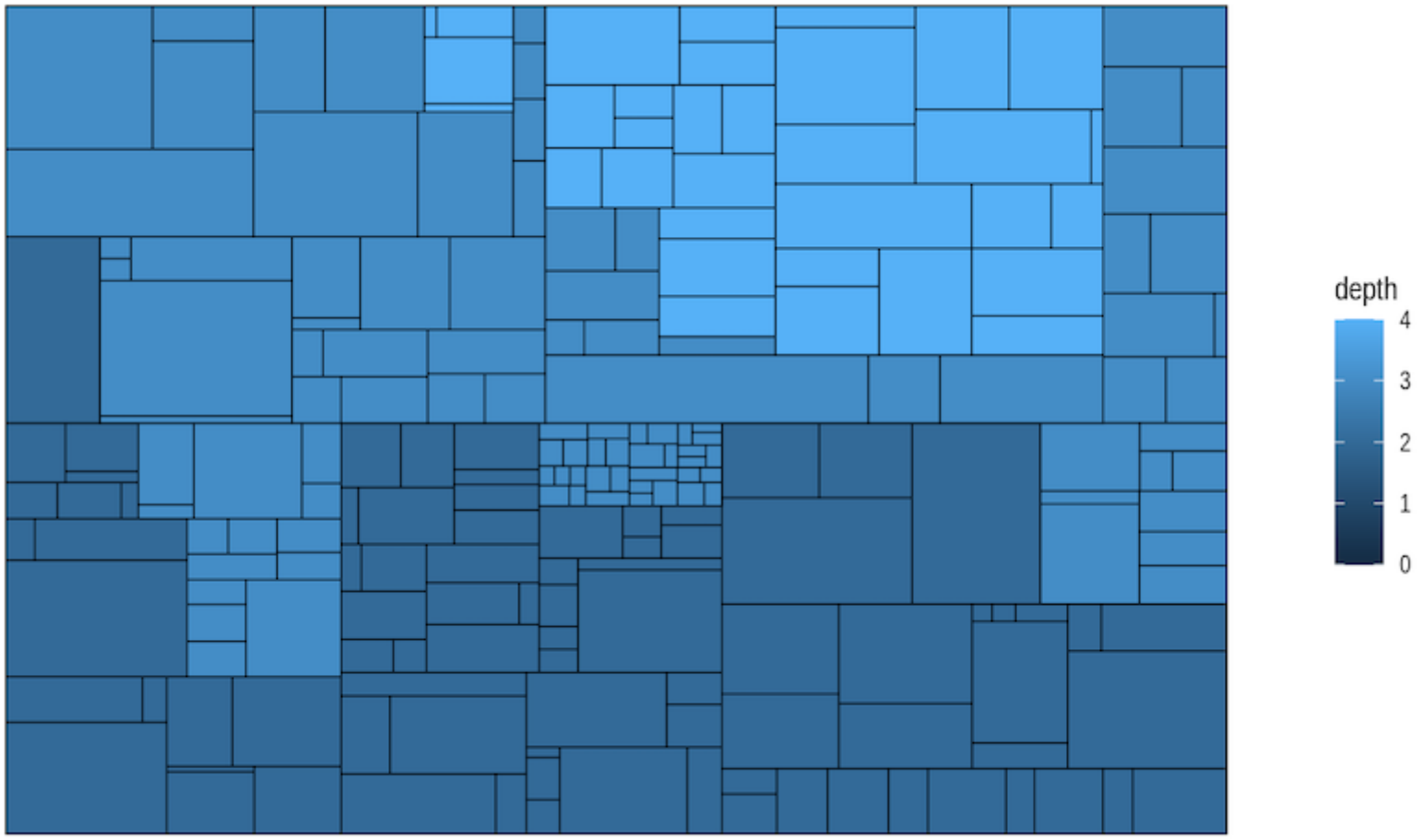
ggraph



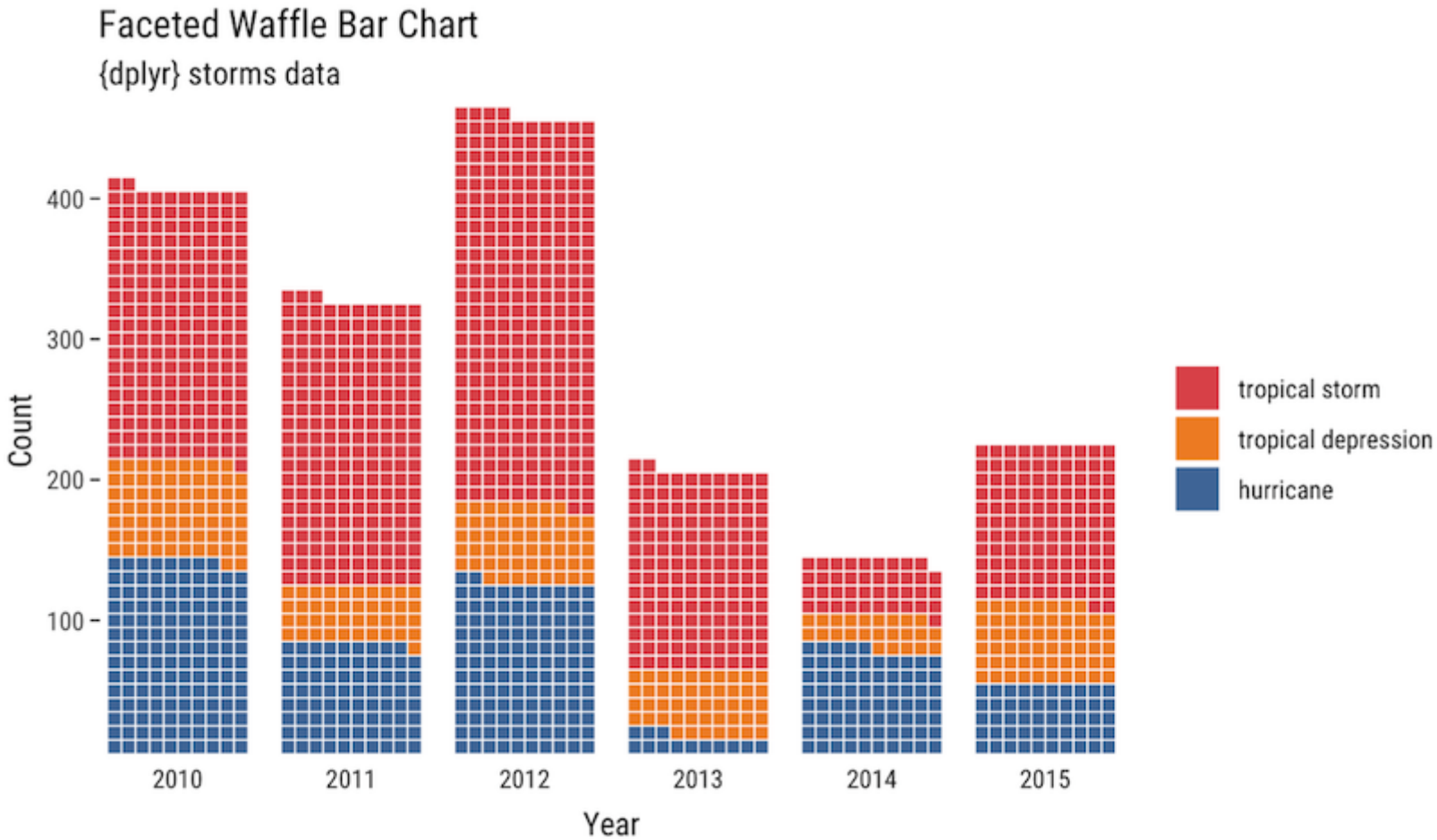
ggraph

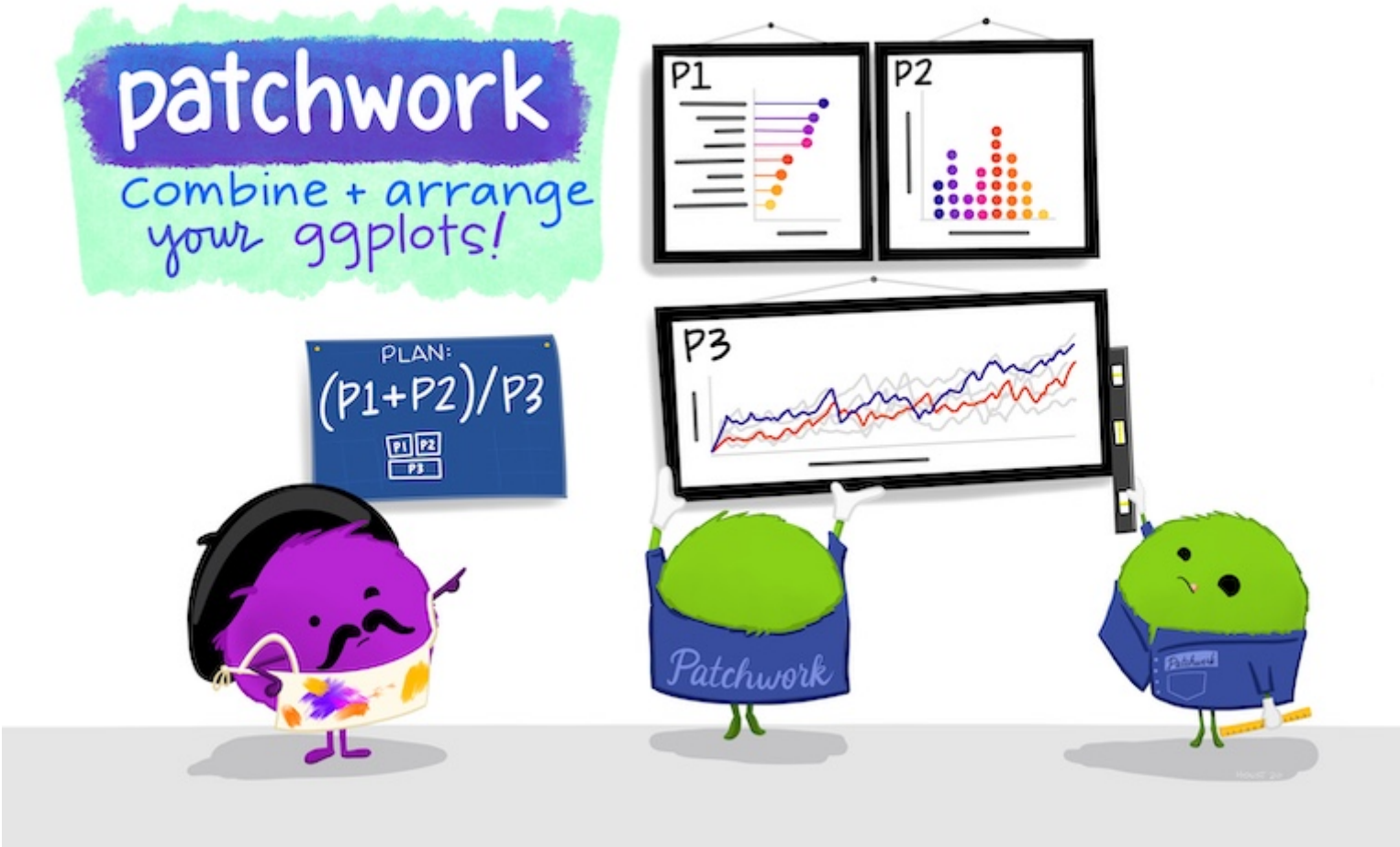


ggraph

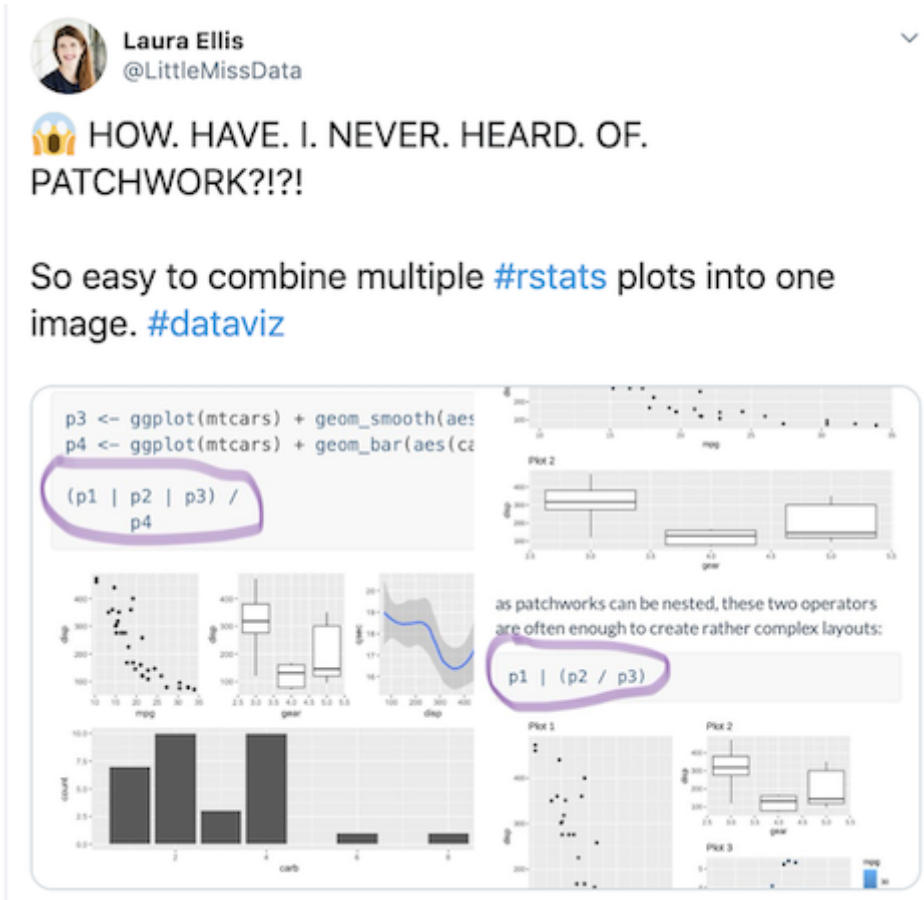


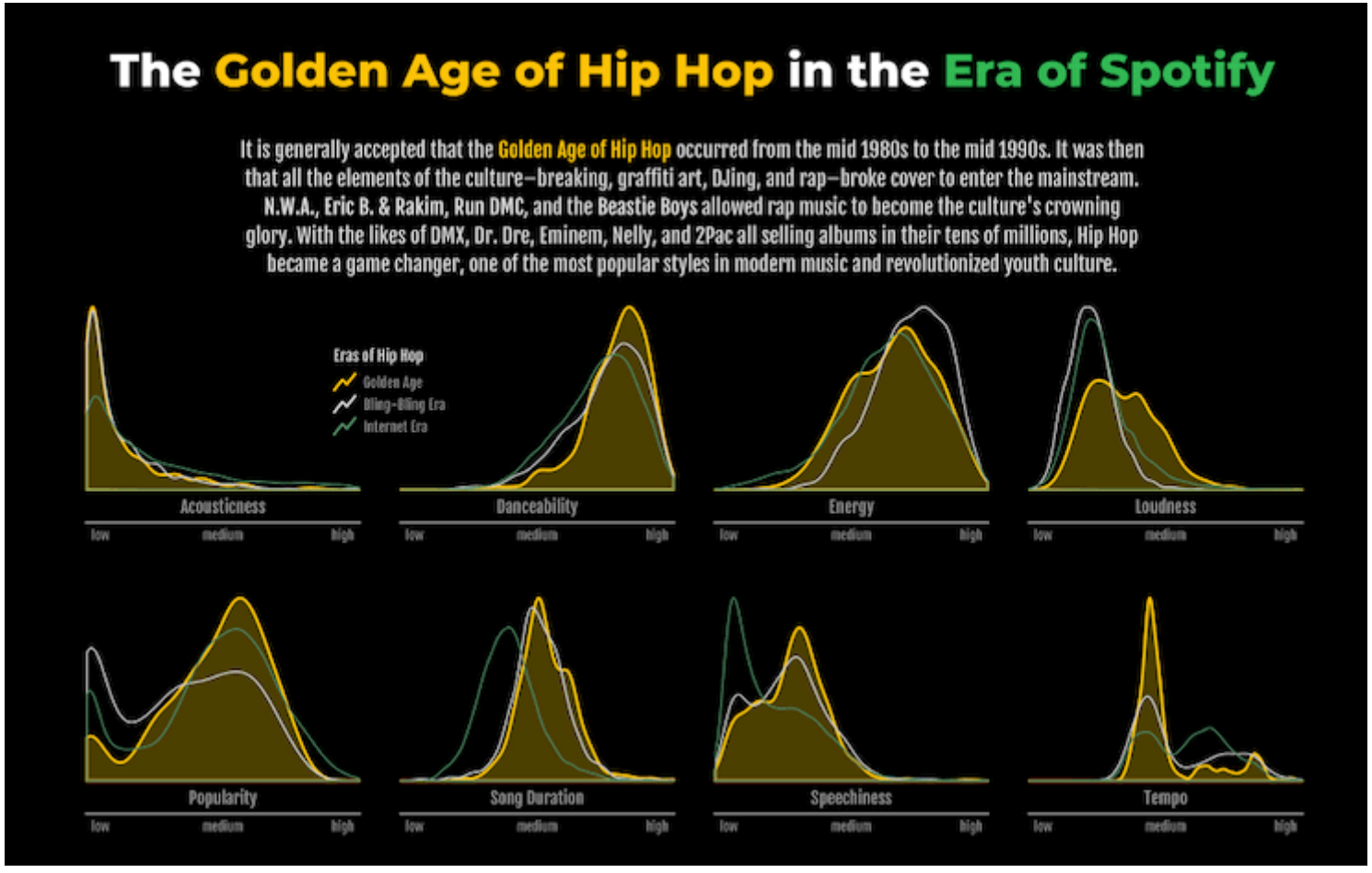
waffle


Faceted Waffle Bar Chart
{dplyr} storms data
Count
400
300
200
100
2010
2011
2012
2013
2014
2015
Year
tropical storm
tropical depression
hurricane

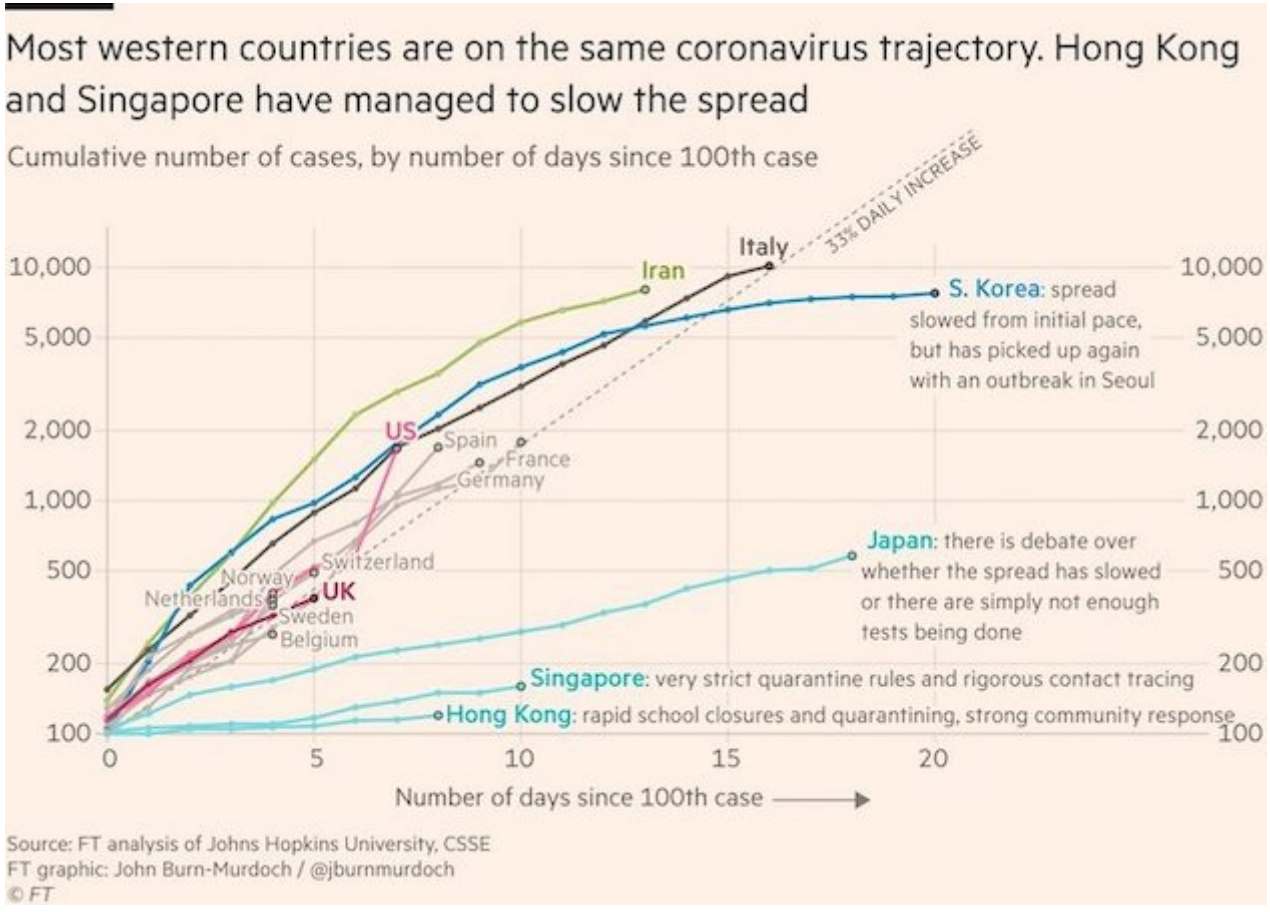

patchwork
combine + arrange your ggplots!
PLAN:
(P1+P2)/P3
P1
P2
P3
Patchwork

patchwork


Laura Ellis
@LittleMissData
HOW. HAVE. I. NEVER. HEARD. OF. PATCHWORK?!?!
So easy to combine multiple #rstats plots into one image. #dataviz
p3 <- ggplot(mtcars) + geom_smooth(aes
p4 <- ggplot(mtcars) + geom_bar(aes(ca
(p1 | p2 | p3) /
p4
as patchworks can be nested, these two operators are often enough to create rather complex layouts:
p1 | (p2 / p3)


The Golden Age of Hip Hop in the Era of Spotify
It is generally accepted that the Golden Age of Hip Hop occurred from the mid 1980s to the mid 1990s. It was then that all the elements of the culture–breaking, graffiti art, DJing, and rap–broke cover to enter the mainstream. N.W.A., Eric B. & Rakim, Run DMC, and the Beastie Boys allowed rap music to become the culture's crowning glory. With the likes of DMX, Dr. Dre, Eminem, Nelly, and 2Pac all selling albums in their tens of millions, Hip Hop became a game changer, one of the most popular styles in modern music and revolutionized youth culture.
Eras of Hip Hop
Golden Age
Bling-Bling Era
Internet Era
Acousticness
Danceability
Energy
Loudness
Popularity
Song Duration
Speechiness
Tempo
low
medium
high

shadowtext


Most western countries are on the same coronavirus trajectory. Hong Kong and Singapore have managed to slow the spread
Cumulative number of cases, by number of days since 100th case
33% DAILY INCREASE
Italy
Iran
S. Korea: spread slowed from initial pace, but has picked up again with an outbreak in Seoul
US
Spain
France
Germany
Switzerland
Norway
Netherlands
UK
Sweden
Belgium
Japan: there is debate over whether the spread has slowed or there are simply not enough tests being done
Singapore: very strict quarantine rules and rigorous contact tracing
Hong Kong: rapid school closures and quarantining, strong community response
10,000
5,000
2,000
1,000
500
200
100
0
5
10
15
20
Number of days since 100th case
Source: FT analysis of Johns Hopkins University, CSSE
FT graphic: John Burn-Murdoch / @jburnmurdoch
© FT

Your Turn

1. Use one of the packages above to make a unique plot. For example, you might use dumbell plots in the `ggalt` package to show change in the Hispanic/Latino population from 2017-2018 to 2018-2019 for all districts.

2. When you finish your plot, email it to me at david@rfortherestofus.com! I'd love to see what you come up with.