Data Visualization. 9 - Case Studies

Kieran Healy
Code Horizons

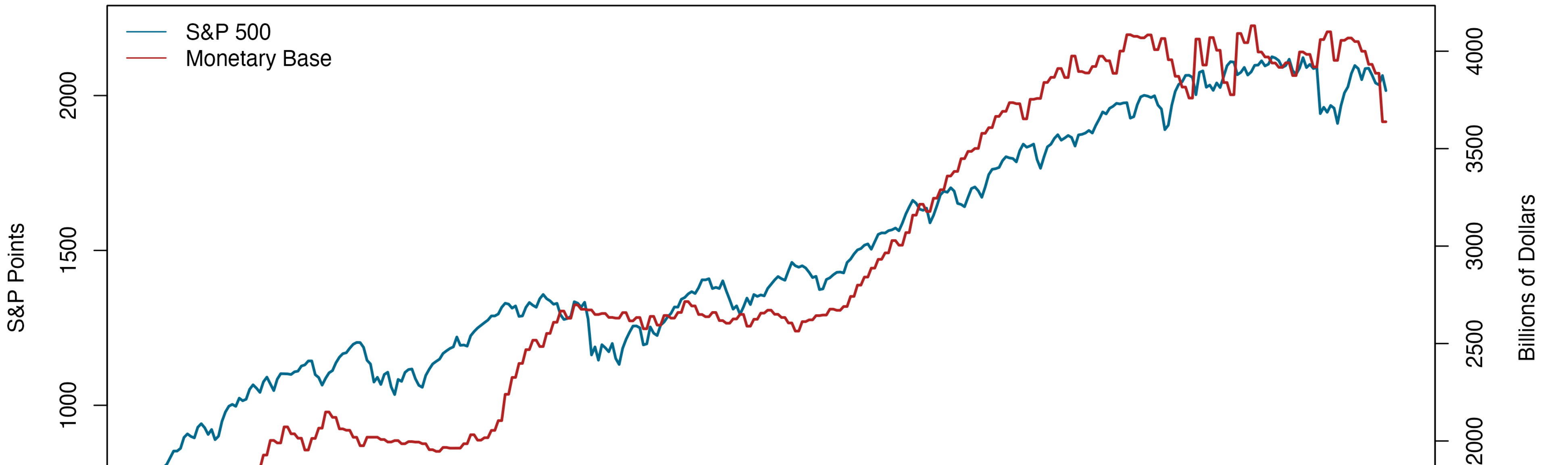
December 10, 2023

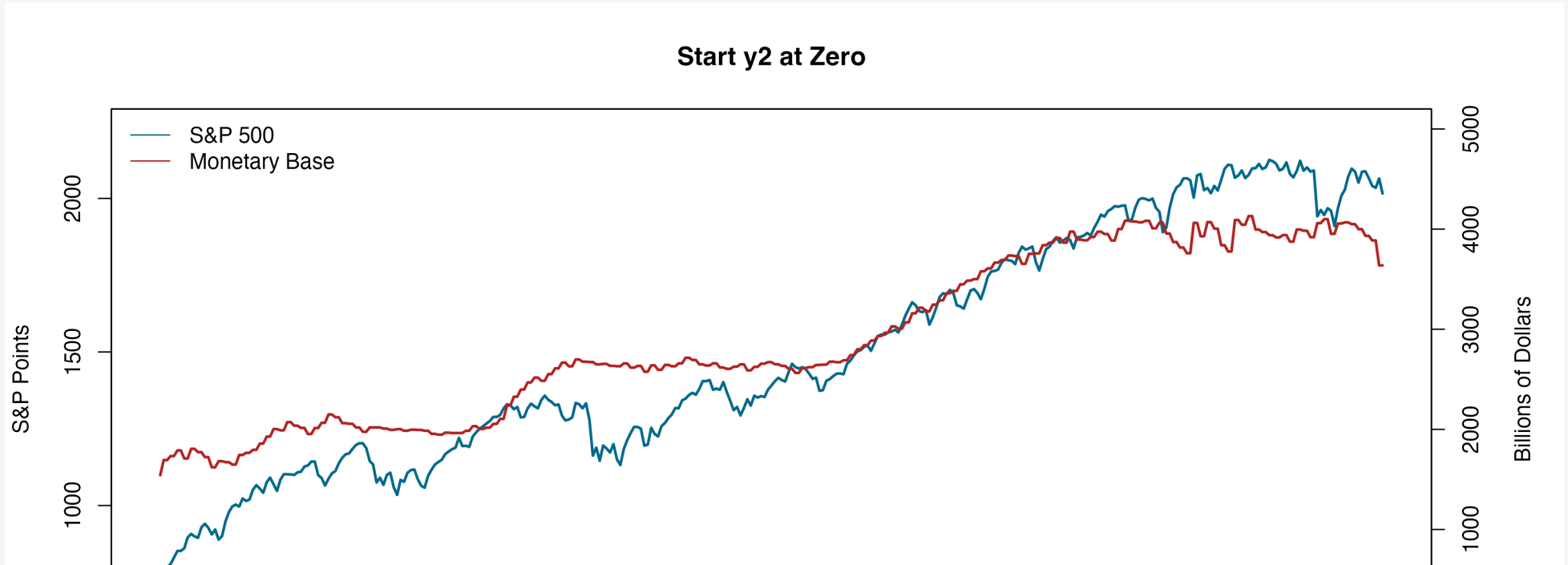
Some Case Studies

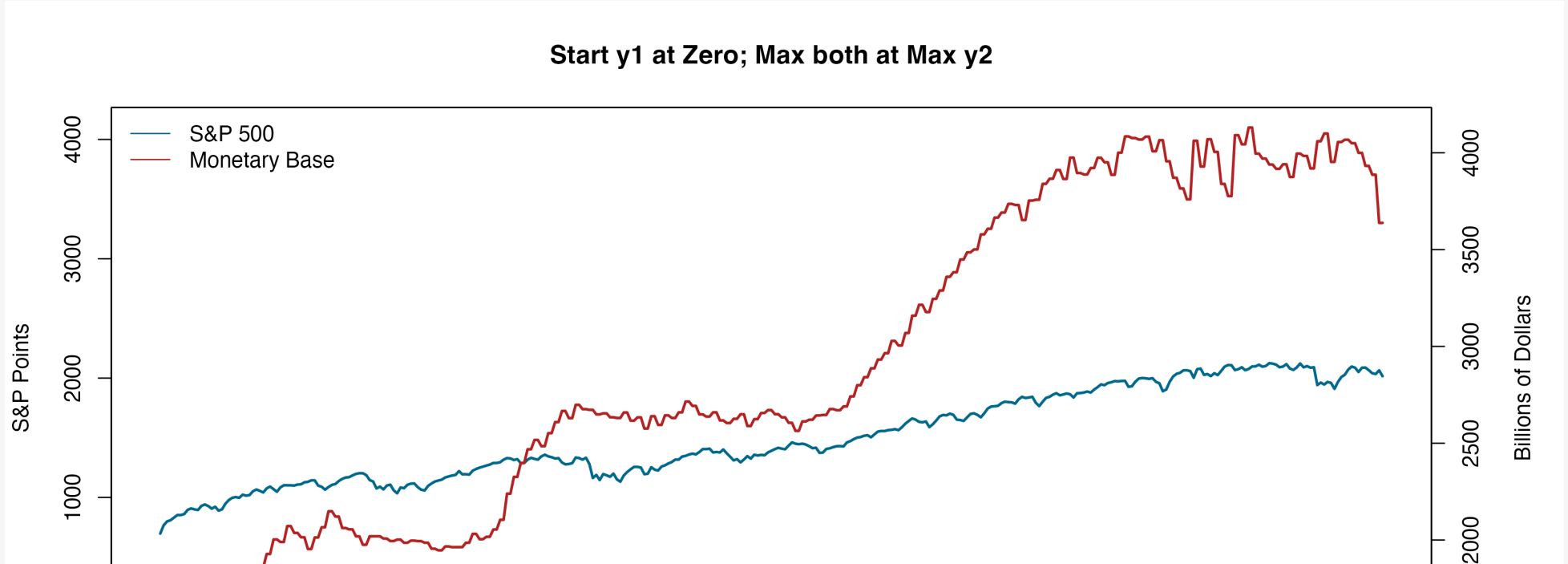
Load our packages

```
library(here)       # manage file paths
library(tidyverse)  # your friend and mine
library(cavax)      # california vaccination exemption data
library(colorspace) # luminance-balanced palettes
library(demog)      # demographic data for a graph
library(ggforce)    # useful enhancements to ggplot
library(ggrepel)    # Text and labels
library(gssr)       # the gss packaged for r
library(patchwork)  # compose multiple plots

library(scales)     # scale adjustments and enhancements
library(socviz)     # data and some useful functions
```

Two y-axes


S&P 500
Monetary Base
S&P Points
2000
1500
1000
Billions of Dollars
4000
3500
3000
2500
2000


Start y2 at Zero
S&P 500
Monetary Base
S&P Points
1000
1500
2000
Billions of Dollars
1000
2000
3000
4000
5000


Start y1 at Zero; Max both at Max y2
S&P 500
Monetary Base
S&P Points
Billions of Dollars
1000
2000
3000
4000
2000
2500
3000
3500
4000

What to do instead?

```
fredts ← as_tibble(fredts)
fredts
```

```
# A tibble: 357 × 5
   date       sp500 monbase sp500_i monbase_i
   <date>     <dbl>   <int>   <dbl>     <dbl>
 1 2009-03-11  697. 1542228    100       100
 2 2009-03-18  767. 1693133    110.      110.
 3 2009-03-25  799. 1693133    115.      110.
 4 2009-04-01  809. 1733017    116.      112.
 5 2009-04-08  831. 1733017    119.      112.
 6 2009-04-15  852. 1789878    122.      116.
 7 2009-04-22  852. 1789878    122.      116.
 8 2009-04-29  861. 1709369    124.      111.
 9 2009-05-06  896. 1709369    129.      111.
10 2009-05-13  908. 1805373    130.      117.
# ℹ 347 more rows
```

Pivot the data

```
1  fredts
```

```
# A tibble: 357 × 5
   date       sp500 monbase sp500_i monbase_i
   <date>     <dbl>   <int>   <dbl>     <dbl>
 1 2009-03-11  697. 1542228    100        100
 2 2009-03-18  767. 1693133    110.       110.
 3 2009-03-25  799. 1693133    115.       110.
 4 2009-04-01  809. 1733017    116.       112.
 5 2009-04-08  831. 1733017    119.       112.
 6 2009-04-15  852. 1789878    122.       116.
 7 2009-04-22  852. 1789878    122.       116.
 8 2009-04-29  861. 1709369    124.       111.
 9 2009-05-06  896. 1709369    129.       111.
10 2009-05-13  908. 1805373    130.       117.
# ℹ 347 more rows
```

Pivot the data

```
1  fredts ▷
2    select(date, sp500_i, monbase_i)
```

```
# A tibble: 357 × 3
   date       sp500_i monbase_i
   <date>       <dbl>     <dbl>
 1 2009-03-11    100       100
 2 2009-03-18    110.      110.
 3 2009-03-25    115.      110.
 4 2009-04-01    116.      112.
 5 2009-04-08    119.      112.
 6 2009-04-15    122.      116.
 7 2009-04-22    122.      116.
 8 2009-04-29    124.      111.
 9 2009-05-06    129.      111.
10 2009-05-13    130.      117.
# ℹ 347 more rows
```

Pivot the data

```
fredts ▷
  select(date, sp500_i, monbase_i)  ▷
  pivot_longer(sp500_i:monbase_i,
               names_to = "series",
               values_to = "score")
```

```
# A tibble: 714 × 3
   date       series    score
   <date>     <chr>     <dbl>
 1 2009-03-11 sp500_i    100
 2 2009-03-11 monbase_i  100
 3 2009-03-18 sp500_i    110.
 4 2009-03-18 monbase_i  110.
 5 2009-03-25 sp500_i    115.
 6 2009-03-25 monbase_i  110.
 7 2009-04-01 sp500_i    116.
 8 2009-04-01 monbase_i  112.
 9 2009-04-08 sp500_i    119.
10 2009-04-08 monbase_i  112.
# ℹ 704 more rows
```

Pivot the data

```
1 fredts ▷
2   select(date, sp500_i, monbase_i)  ▷
3   pivot_longer(sp500_i:monbase_i,
4                names_to = "series",
5                values_to = "score") →
6   fredts_m
```

Pivot the data

```
fredts ▷
  select(date, sp500_i, monbase_i)  ▷
  pivot_longer(sp500_i:monbase_i,
               names_to = "series",
               values_to = "score") →
  fredts_m
```

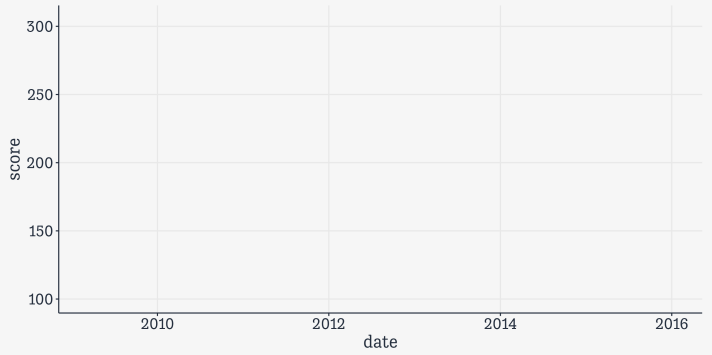
Make two plots

```
1  fredts_m
```

```
# A tibble: 714 × 3
   date       series    score
   <date>     <chr>     <dbl>
 1 2009-03-11 sp500_i    100
 2 2009-03-11 monbase_i  100
 3 2009-03-18 sp500_i    110.
 4 2009-03-18 monbase_i  110.
 5 2009-03-25 sp500_i    115.
 6 2009-03-25 monbase_i  110.
 7 2009-04-01 sp500_i    116.
 8 2009-04-01 monbase_i  112.
 9 2009-04-08 sp500_i    119.
10 2009-04-08 monbase_i  112.
# ℹ 704 more rows
```

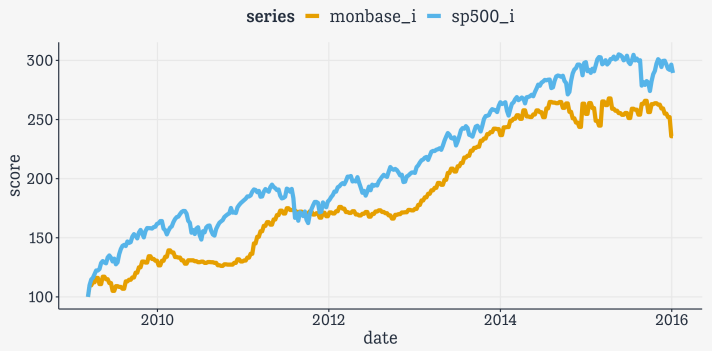
Make two plots

```
fredts_m ▷
  ggplot(mapping =
           aes(x = date,
               y = score,
               color = series))
```



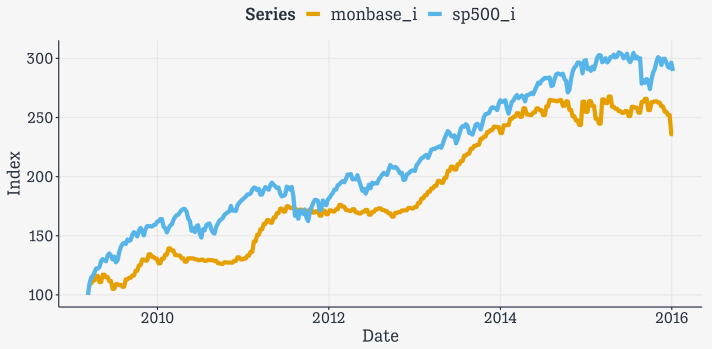
Make two plots

```
fredts_m ▷
  ggplot(mapping =
           aes(x = date,
               y = score,
               color = series)) +
  geom_line(linewidth = 2)
```



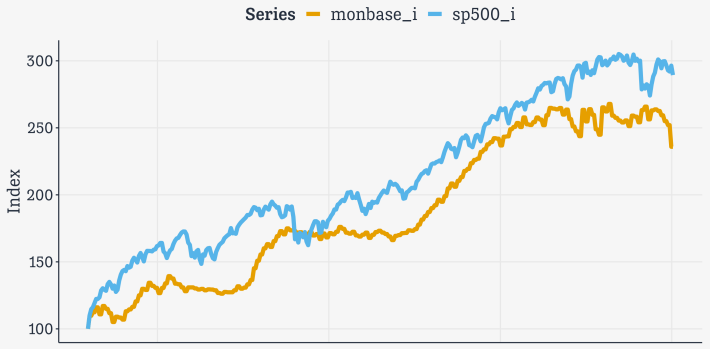
Make two plots

```
fredts_m ▷
  ggplot(mapping =
           aes(x = date,
               y = score,
               color = series)) +
  geom_line(linewidth = 2) +
  labs(x = "Date", y = "Index",
       color = "Series")
```



Make two plots

```
fredts_m ▷
  ggplot(mapping =
           aes(x = date,
               y = score,
               color = series)) +
  geom_line(linewidth = 2) +
  labs(x = "Date", y = "Index",
       color = "Series") +
  theme(axis.title.x = element_blank(),
       axis.text.x = element_blank(),
       axis.ticks.x = element_blank())
```



Make two plots

```
fredts_m ▷
  ggplot(mapping =
           aes(x = date,
               y = score,
               color = series)) +
  geom_line(linewidth = 2) +
  labs(x = "Date", y = "Index",
       color = "Series") +
  theme(axis.title.x = element_blank(),
       axis.text.x = element_blank(),
       axis.ticks.x = element_blank()) →
  p1
```

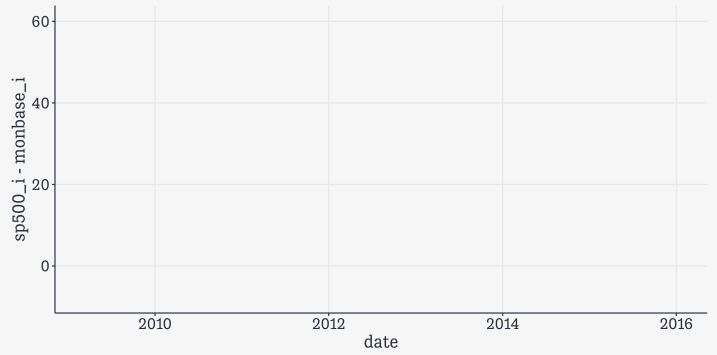
Make two plots

```
fredts_m ▷
  ggplot(mapping =
           aes(x = date,
               y = score,
               color = series)) +
  geom_line(linewidth = 2) +
  labs(x = "Date", y = "Index",
       color = "Series") +
  theme(axis.title.x = element_blank(),
       axis.text.x = element_blank(),
       axis.ticks.x = element_blank()) →
  p1

# The original df
fredts
```

```
# A tibble: 357 × 5
   date       sp500 monbase sp500_i monbase_i
   <date>     <dbl>   <int>   <dbl>     <dbl>
 1 2009-03-11  697. 1542228    100       100
 2 2009-03-18  767. 1693133    110.      110.
 3 2009-03-25  799. 1693133    115.      110.
 4 2009-04-01  809. 1733017    116.      112.
 5 2009-04-08  831. 1733017    119.      112.
 6 2009-04-15  852. 1789878    122.      116.
 7 2009-04-22  852. 1789878    122.      116.
 8 2009-04-29  861. 1709369    124.      111.
 9 2009-05-06  896. 1709369    129.      111.
10 2009-05-13  908. 1805373    130.      117.
# ℹ 347 more rows
```

Make two plots

```
fredts_m ▷
  ggplot(mapping =
           aes(x = date,
               y = score,
               color = series)) +
  geom_line(linewidth = 2) +
  labs(x = "Date", y = "Index",
       color = "Series") +
  theme(axis.title.x = element_blank(),
       axis.text.x = element_blank(),
       axis.ticks.x = element_blank()) →
  p1

# The original df
fredts ▷
  ggplot(mapping =
          aes(x = date,
              y = sp500_i - monbase_i))
```



Make two plots

```
fredts_m ▷
  ggplot(mapping =
           aes(x = date,
               y = score,
               color = series)) +
  geom_line(linewidth = 2) +
  labs(x = "Date", y = "Index",
       color = "Series") +
  theme(axis.title.x = element_blank(),
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank()) →
  p1

# The original df
fredts ▷
  ggplot(mapping =
           aes(x = date,
               y = sp500_i - monbase_i)) +
  geom_line(linewidth = 1.5)
```

Make two plots

```
fredts_m ▷
  ggplot(mapping =
           aes(x = date,
               y = score,
               color = series)) +
  geom_line(linewidth = 2) +
  labs(x = "Date", y = "Index",
       color = "Series") +
  theme(axis.title.x = element_blank(),
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank()) →
  p1

# The original df
fredts ▷
  ggplot(mapping =
           aes(x = date,
               y = sp500_i - monbase_i)) +
  geom_line(linewidth = 1.5) +
  labs(x = "Date", y = "Difference")
```

Make two plots

```
fredts_m ▷
  ggplot(mapping =
           aes(x = date,
               y = score,
               color = series)) +
  geom_line(linewidth = 2) +
  labs(x = "Date", y = "Index",
       color = "Series") +
  theme(axis.title.x = element_blank(),
       axis.text.x = element_blank(),
       axis.ticks.x = element_blank()) →
  p1

# The original df
fredts ▷
  ggplot(mapping =
           aes(x = date,
               y = sp500_i - monbase_i)) +
  geom_line(linewidth = 1.5) +
  labs(x = "Date", y = "Difference") →
  p2
```

Make two plots

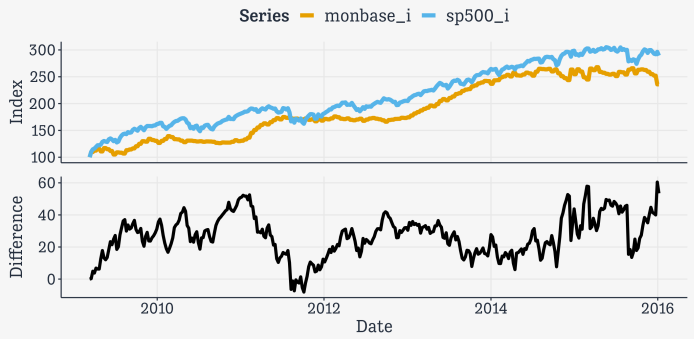
```
fredts_m |>
  ggplot(mapping =
           aes(x = date,
               y = score,
               color = series)) +
  geom_line(linewidth = 2) +
  labs(x = "Date", y = "Index",
       color = "Series") +
  theme(axis.title.x = element_blank(),
       axis.text.x = element_blank(),
       axis.ticks.x = element_blank()) ->
  p1

# The original df
fredts |>
  ggplot(mapping =
           aes(x = date,
               y = sp500_i - monbase_i)) +
  geom_line(linewidth = 1.5) +
  labs(x = "Date", y = "Difference") ->
  p2
```

Combine with patchwork

```
1  library(patchwork)
```

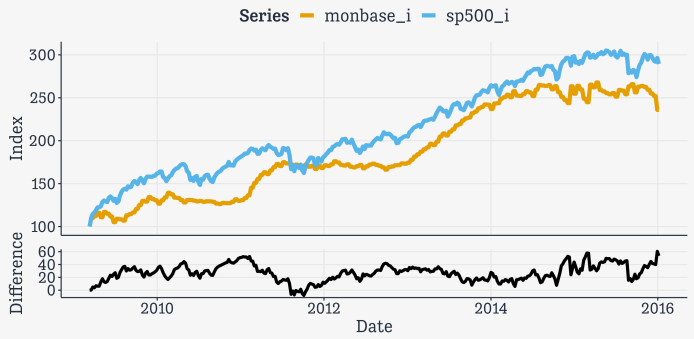
Combine with patchwork

```
1  library(patchwork)
2
3  (p1 / p2)
```



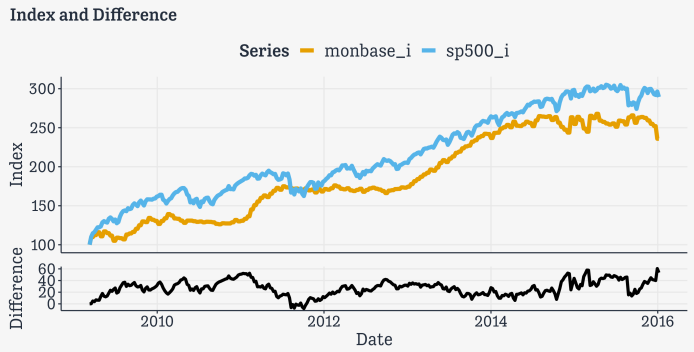
Combine with patchwork

```
1  library(patchwork)
2
3  (p1 / p2) +
4    plot_layout(heights = c(4, 1))
```



Combine with patchwork

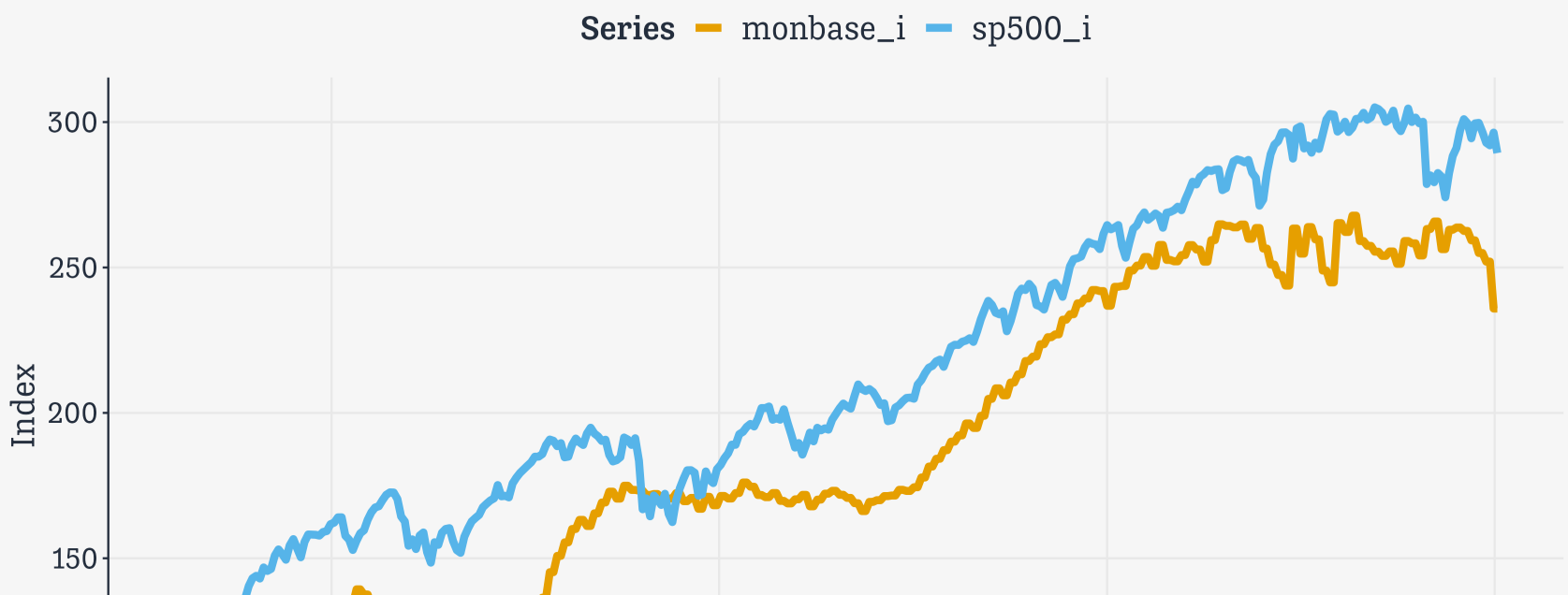
```
1  library(patchwork)
2
3  (p1 / p2) +
4    plot_layout(heights = c(4, 1)) +
5    plot_annotation(title = "Index and Difference")
```



Combine with patchwork

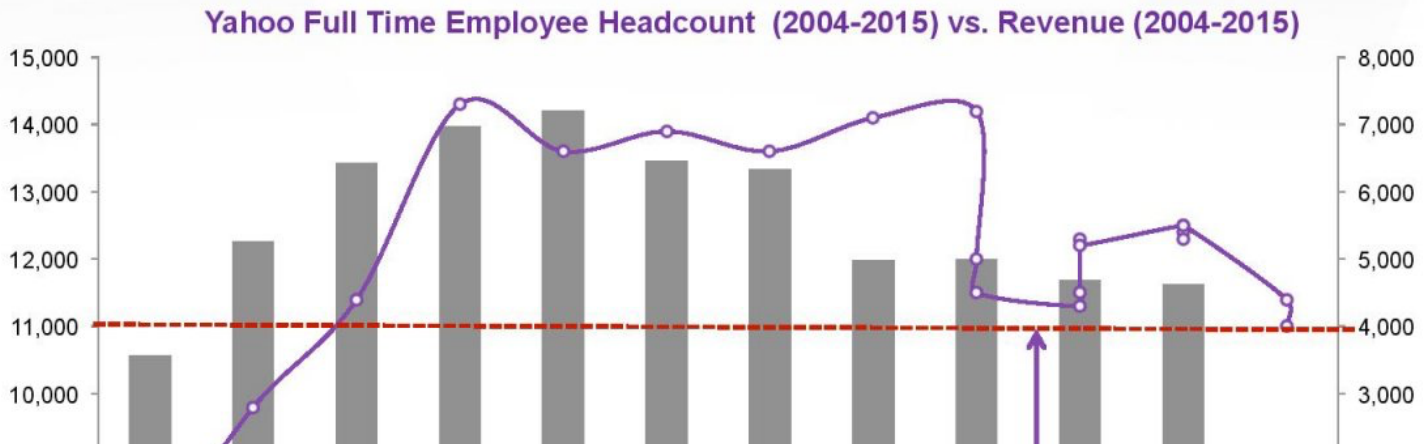
```
1  library(patchwork)
2
3  (p1 / p2) +
4    plot_layout(heights = c(4, 1)) +
5    plot_annotation(title = "Index and Difference") &
6    p_patch
```

Index and Difference



Redrawing a bad slide

Yahoo's Headcount Still Excessively High Given Revenues:



The data

```
yahoo
```

```
# A tibble: 12 × 4
    Year Revenue Employees Mayer
   <dbl>   <dbl>     <dbl> <chr>
 1  2004    3574      7600 No
 2  2005    5257      9800 No
 3  2006    6425     11400 No
 4  2007    6969     14300 No
 5  2008    7208     13600 No
 6  2009    6460     13900 No
 7  2010    6324     13600 No
 8  2011    4984     14100 No
 9  2012    4986     12000 No
10  2012    4986     11500 Yes
11  2013    4680     12200 Yes
12  2014    4618     12500 Yes
```

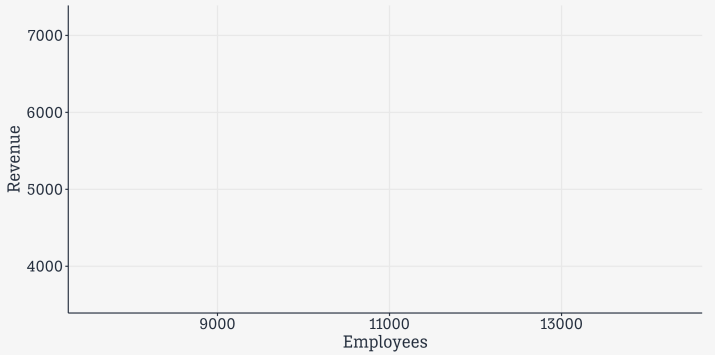
Option 1

```
1  yahoo
```

```
# A tibble: 12 × 4
    Year Revenue Employees Mayer
   <dbl>   <dbl>     <dbl> <chr>
 1  2004    3574      7600 No
 2  2005    5257      9800 No
 3  2006    6425     11400 No
 4  2007    6969     14300 No
 5  2008    7208     13600 No
 6  2009    6460     13900 No
 7  2010    6324     13600 No
 8  2011    4984     14100 No
 9  2012    4986     12000 No
10  2012    4986     11500 Yes
11  2013    4680     12200 Yes
12  2014    4618     12500 Yes
```

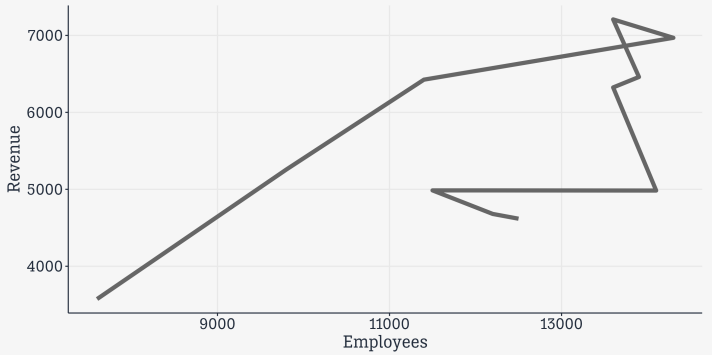
Option 1

```
yahoo ▷
  ggplot(mapping =
           aes(x = Employees,
               y = Revenue))
```



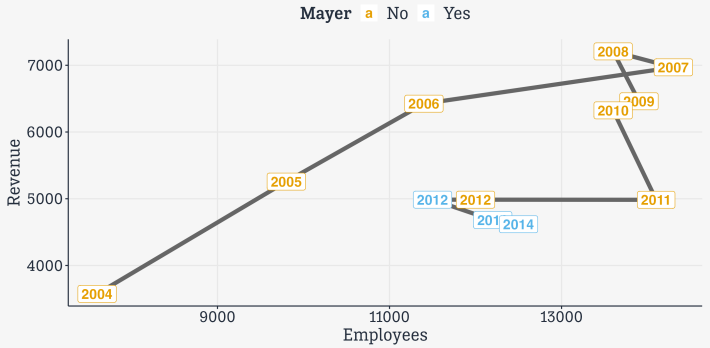
Option 1

```
yahoo ▷
  ggplot(mapping =
           aes(x = Employees,
               y = Revenue)) +
  geom_path(color = "gray40",
            linewidth = rel(2))
```



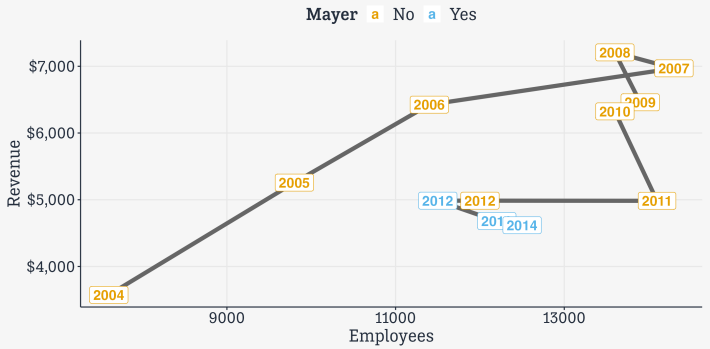
Option 1

```
yahoo ▷
  ggplot(mapping =
           aes(x = Employees,
               y = Revenue)) +
  geom_path(color = "gray40",
            linewidth = rel(2)) +
  geom_label(aes(color = Mayer,
                 label = Year),
             size = rel(5),
             fontface = "bold")
```



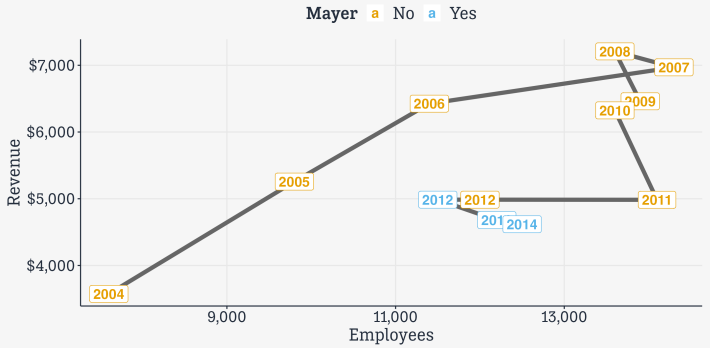
Option 1

```
yahoo ▷
  ggplot(mapping =
           aes(x = Employees,
               y = Revenue)) +
  geom_path(color = "gray40",
            linewidth = rel(2)) +
  geom_label(aes(color = Mayer,
                 label = Year),
             size = rel(5),
             fontface = "bold") +
  scale_y_continuous(labels = label_dollar())
```



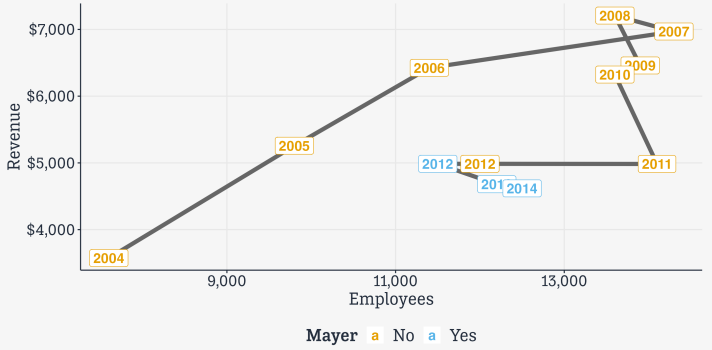
Option 1

```
yahoo ▷
  ggplot(mapping =
           aes(x = Employees,
               y = Revenue)) +
  geom_path(color = "gray40",
            linewidth = rel(2)) +
  geom_label(aes(color = Mayer,
                 label = Year),
             size = rel(5),
             fontface = "bold") +
  scale_y_continuous(labels = label_dollar()) +
  scale_x_continuous(labels = label_comma())
```



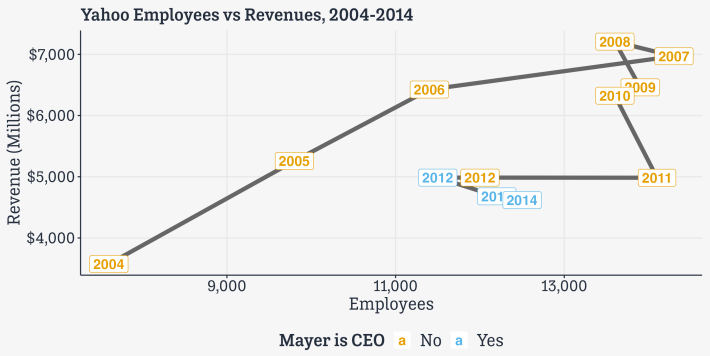
Option 1

```
yahoo ▷
  ggplot(mapping =
           aes(x = Employees,
               y = Revenue)) +
  geom_path(color = "gray40",
            linewidth = rel(2)) +
  geom_label(aes(color = Mayer,
                 label = Year),
             size = rel(5),
             fontface = "bold") +
  scale_y_continuous(labels = label_dollar()) +
  scale_x_continuous(labels = label_comma()) +
  theme(legend.position = "bottom")
```



Option 1

```
yahoo ▷
  ggplot(mapping =
           aes(x = Employees,
               y = Revenue)) +
  geom_path(color = "gray40",
            linewidth = rel(2)) +
  geom_label(aes(color = Mayer,
                 label = Year),
             size = rel(5),
             fontface = "bold") +
  scale_y_continuous(labels = label_dollar()) +
  scale_x_continuous(labels = label_comma()) +
  theme(legend.position = "bottom") +
  labs(color = "Mayer is CEO",
       x = "Employees", y = "Revenue (Millions)",
       title = "Yahoo Employees vs Revenues, 2004-2014")
```

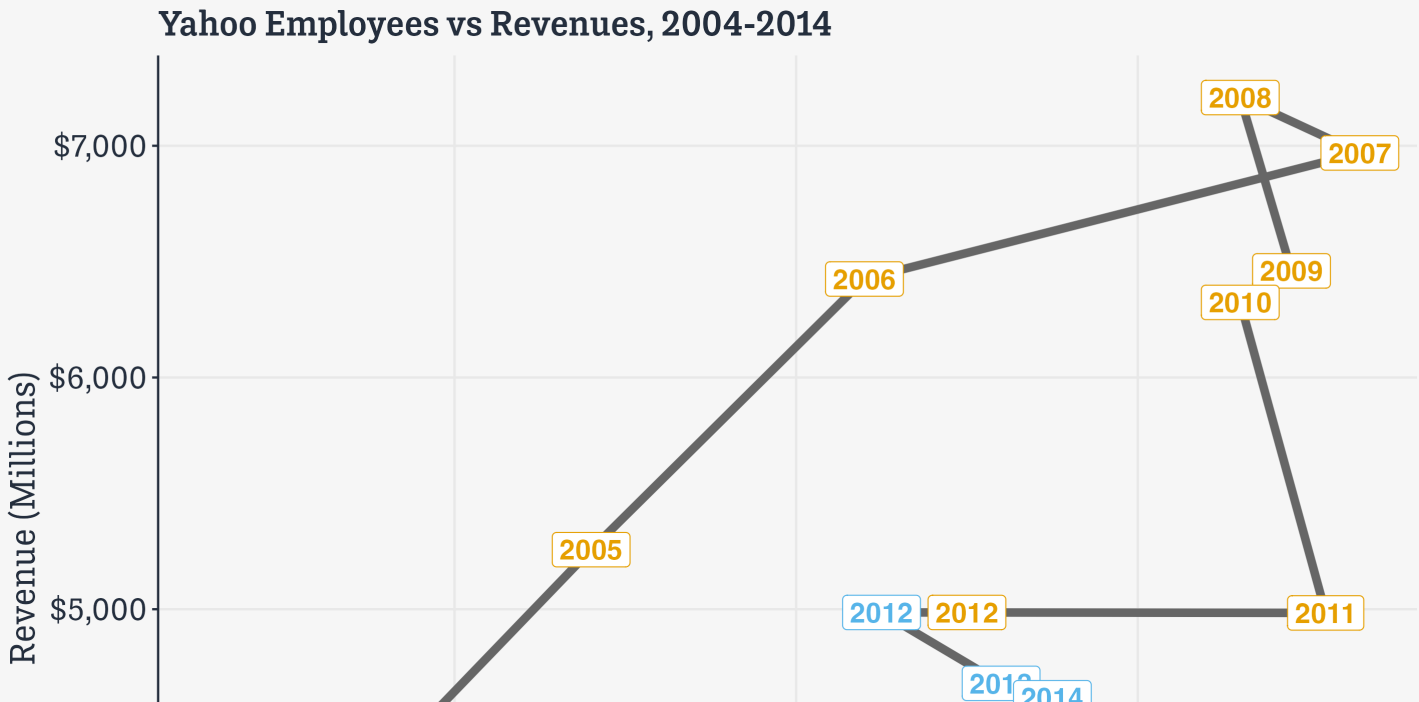


Option 1

```
yahoo ▷
  ggplot(mapping =
           aes(x = Employees,
               y = Revenue)) +
  geom_path(color = "gray40",
            linewidth = rel(2)) +
  geom_label(aes(color = Mayer,
                 label = Year),
             size = rel(5),
             fontface = "bold") +
  scale_y_continuous(labels = label_dollar()) +
  scale_x_continuous(labels = label_comma()) +
  theme(legend.position = "bottom") +
  labs(color = "Mayer is CEO",
       x = "Employees", y = "Revenue (Millions)",
       title = "Yahoo Employees vs Revenues, 2004-2014") →
  yahoo1
```

Option 1

```
yahoo ▷
  ggplot(mapping =
           aes(x = Employees,
               y = Revenue)) +
  geom_path(color = "gray40",
            linewidth = rel(2)) +
  geom_label(aes(color = Mayer,
                 label = Year),
             size = rel(5),
             fontface = "bold") +
  scale_y_continuous(labels = label_dollar()) +
  scale_x_continuous(labels = label_comma()) +
  theme(legend.position = "bottom") +
  labs(color = "Mayer is CEO",
       x = "Employees", y = "Revenue (Millions)",
       title = "Yahoo Employees vs Revenues, 2004-2014") →
  yahoo1
```


Yahoo Employees vs Revenues, 2004-2014
Revenue (Millions)
$7,000
$6,000
$5,000
2008
2007
2006
2009
2010
2005
2012
2012
2011
2014

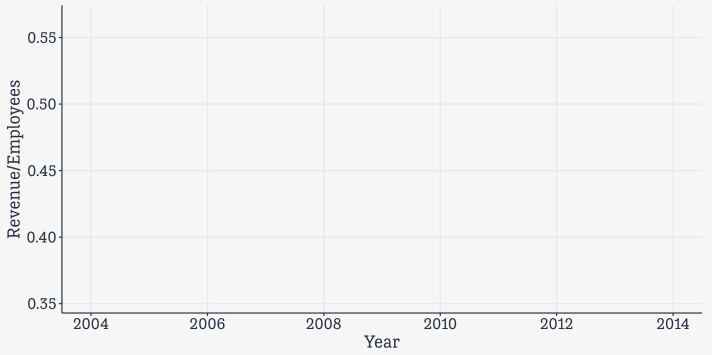
Alternatively ...

```
1  yahoo
```

```
# A tibble: 12 × 4
    Year Revenue Employees Mayer
   <dbl>   <dbl>     <dbl> <chr>
 1  2004    3574      7600 No
 2  2005    5257      9800 No
 3  2006    6425     11400 No
 4  2007    6969     14300 No
 5  2008    7208     13600 No
 6  2009    6460     13900 No
 7  2010    6324     13600 No
 8  2011    4984     14100 No
 9  2012    4986     12000 No
10  2012    4986     11500 Yes
11  2013    4680     12200 Yes
12  2014    4618     12500 Yes
```

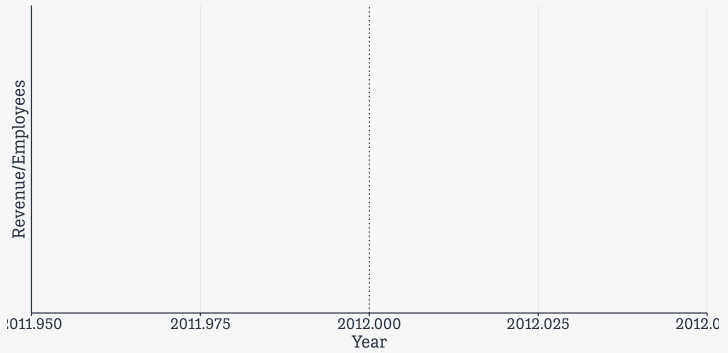
Alternatively ...

```
yahoo ▷
  ggplot(mapping =
           aes(x = Year,
               y = Revenue/Employees))
```



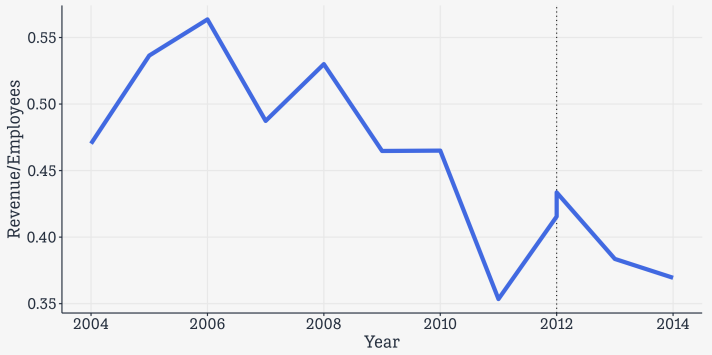
Alternatively ...

```
yahoo ▷
  ggplot(mapping =
           aes(x = Year,
               y = Revenue/Employees)) +
  geom_vline(xintercept = 2012,
             linewidth = rel(0.5),
             linetype = "dotted")
```



Alternatively ...

```
yahoo ▷
  ggplot(mapping =
           aes(x = Year,
               y = Revenue/Employees)) +
  geom_vline(xintercept = 2012,
             linewidth = rel(0.5),
             linetype = "dotted") +
  geom_line(color = "royalblue", linewidth = rel(2))
```

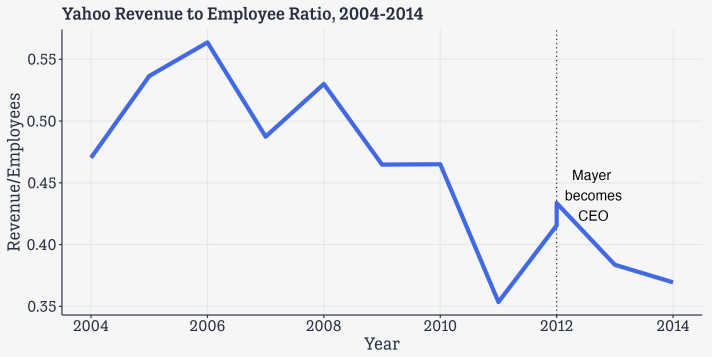


Alternatively ...

```
yahoo ▷
  ggplot(mapping =
           aes(x = Year,
               y = Revenue/Employees)) +
  geom_vline(xintercept = 2012,
             linewidth = rel(0.5),
             linetype = "dotted") +
  geom_line(color = "royalblue", linewidth = rel(2)) +
  annotate("text", x = 2012.6, y = 0.44,
           label = "Mayer\n becomes\n CEO", size = rel(5))
```

Alternatively ...

```
yahoo ▷
  ggplot(mapping =
           aes(x = Year,
               y = Revenue/Employees)) +
  geom_vline(xintercept = 2012,
             linewidth = rel(0.5),
             linetype = "dotted") +
  geom_line(color = "royalblue", linewidth = rel(2)) +
  annotate("text", x = 2012.6, y = 0.44,
           label = "Mayer\n becomes\n CEO", size = rel(5)) +
  labs(title = "Yahoo Revenue to Employee Ratio, 2004-2014",
       x = "Year",
       y = "Revenue/Employees")
```



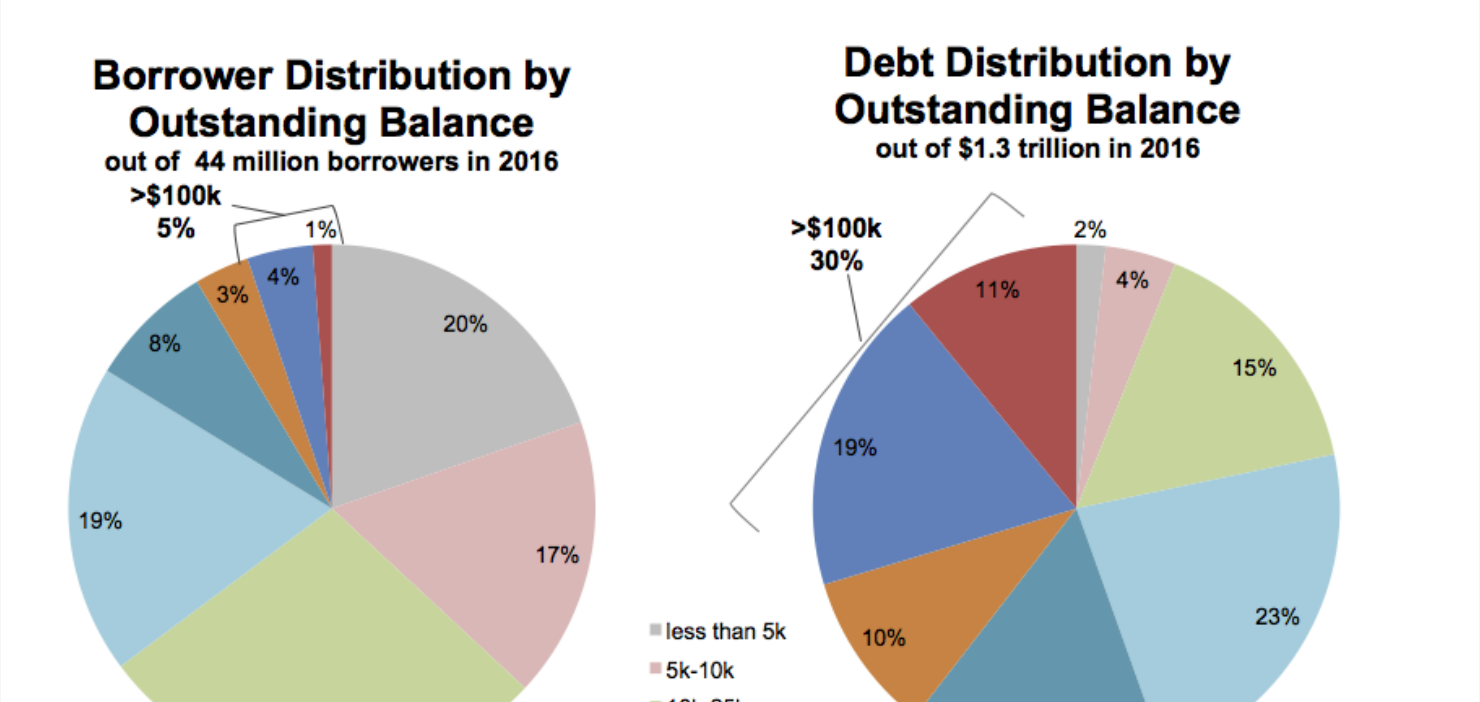
Alternatively ...

```
yahoo ▷
  ggplot(mapping =
           aes(x = Year,
               y = Revenue/Employees)) +
  geom_vline(xintercept = 2012,
             linewidth = rel(0.5),
             linetype = "dotted") +
  geom_line(color = "royalblue", linewidth = rel(2)) +
  annotate("text", x = 2012.6, y = 0.44,
           label = "Mayer\n becomes\n CEO", size = rel(5)) +
  labs(title = "Yahoo Revenue to Employee Ratio, 2004-2014",
       x = "Year",
       y = "Revenue/Employees") →
  yahoo2
```


Yahoo Revenue to Employee Ratio, 2004-2014
0.55
0.50
0.45
Revenue/Employees
Mayer
becomes
CEO

Say no to pie

Pie charts are easy to mess up



The data

```
studebt
```

```
# A tibble: 16 × 4
   Debt      type        pct Debtrc
   <ord>     <fct>     <int> <ord>
 1 Under $5  Borrowers    20 Under $5
 2 $5-$10    Borrowers    17 $5-$10
 3 $10-$25   Borrowers    28 $10-$25
 4 $25-$50   Borrowers    19 $25-$50
 5 $50-$75   Borrowers     8 $50-$75
 6 $75-$100  Borrowers     3 $75-$100
 7 $100-$200 Borrowers     4 $100-$200
 8 Over $200 Borrowers     1 Over $200
 9 Under $5  Balances      2 Under $5
10 $5-$10    Balances      4 $5-$10
11 $10-$25   Balances     15 $10-$25
12 $25-$50   Balances     23 $25-$50
13 $50-$75   Balances     16 $50-$75
14 $75-$100  Balances     10 $75-$100
15 $100-$200 Balances     19 $100-$200
16 Over $200 Balances     11 Over $200
```

`Debt` and `Debtrc` are both ordered factors.

A little prep work

```
p_ylab ← "Amount Owed, in thousands of Dollars"
p_title ← "Outstanding Student Loans"
p_subtitle ← "44 million borrowers owe a total of $1.3 trillion"
p_caption ← "Source: FRB NY"

studebt ← studebt ▷
  mutate(type_label = recode(type, "Borrowers" = "Percent of all Borrowers",
                             "Balances" = "Percent of all Balances"))



studebt
```

```
# A tibble: 16 × 5
   Debt      type        pct Debtrc    type_label
   <ord>     <fct>     <int> <ord>     <fct>
 1 Under $5  Borrowers    20 Under $5  Percent of all Borrowers
 2 $5-$10    Borrowers    17 $5-$10    Percent of all Borrowers
 3 $10-$25   Borrowers    28 $10-$25   Percent of all Borrowers
 4 $25-$50   Borrowers    19 $25-$50   Percent of all Borrowers
 5 $50-$75   Borrowers     8 $50-$75   Percent of all Borrowers
 6 $75-$100  Borrowers     3 $75-$100  Percent of all Borrowers
 7 $100-$200 Borrowers     4 $100-$200 Percent of all Borrowers
 8 Over $200 Borrowers     1 Over $200 Percent of all Borrowers
 9 Under $5  Balances      2 Under $5  Percent of all Balances
10 $5-$10    Balances      4 $5-$10    Percent of all Balances
11 $10-$25   Balances     15 $10-$25   Percent of all Balances
12 $25-$50   Balances     23 $25-$50   Percent of all Balances
```

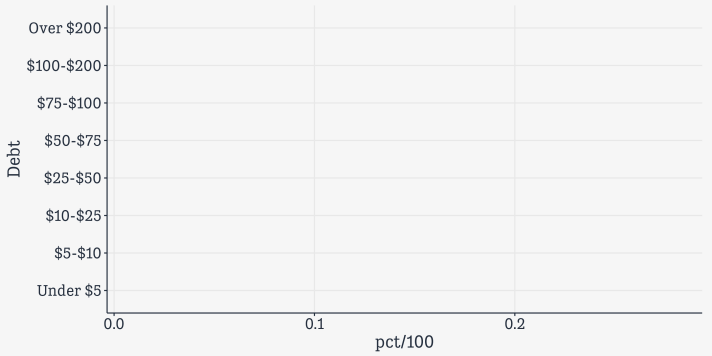
Debt Plot 1

```
1 studebt
```

```
# A tibble: 16 × 5
   Debt      type        pct Debtrc    type_label
   <ord>     <fct>     <int> <ord>     <fct>
 1 Under $5  Borrowers    20 Under $5  Percent of
all Borrowers
 2 $5-$10    Borrowers    17 $5-$10    Percent of
all Borrowers
 3 $10-$25   Borrowers    28 $10-$25   Percent of
all Borrowers
 4 $25-$50   Borrowers    19 $25-$50   Percent of
all Borrowers
 5 $50-$75   Borrowers     8 $50-$75   Percent of
all Borrowers
 6 $75-$100  Borrowers     3 $75-$100  Percent of
all Borrowers
 7 $100-$200 Borrowers     4 $100-$200 Percent of
all Borrowers
 8 Over $200 Borrowers     1 Over $200 Percent of
all Borrowers
 9 Under $5  Balances      2 Under $5  Percent of
all Balances
10 $5-$10    Balances      4 $5-$10    Percent of
all Balances
11 $10-$25   Balances     15 $10-$25   Percent of
all Balances
```

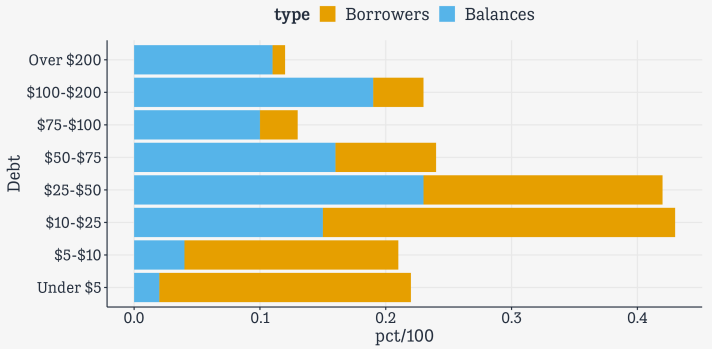
Debt Plot 1

```
studebt ▷
  ggplot(mapping =
           aes(x = pct/100,
               y = Debt,
               fill = type))
```



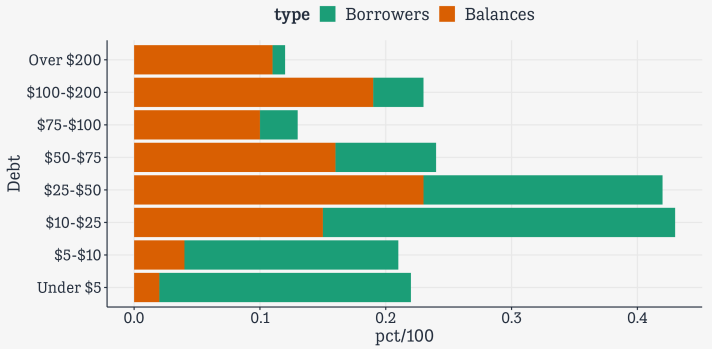
Debt Plot 1

```
studebt ▷
  ggplot(mapping =
           aes(x = pct/100,
               y = Debt,
               fill = type)) +
  geom_col()
```



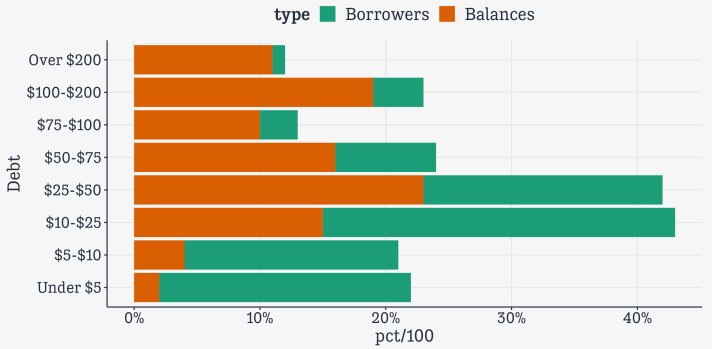
Debt Plot 1

```
studebt ▷
  ggplot(mapping =
           aes(x = pct/100,
               y = Debt,
               fill = type)) +
  geom_col() +
  scale_fill_brewer(type = "qual",
                    palette = "Dark2")
```



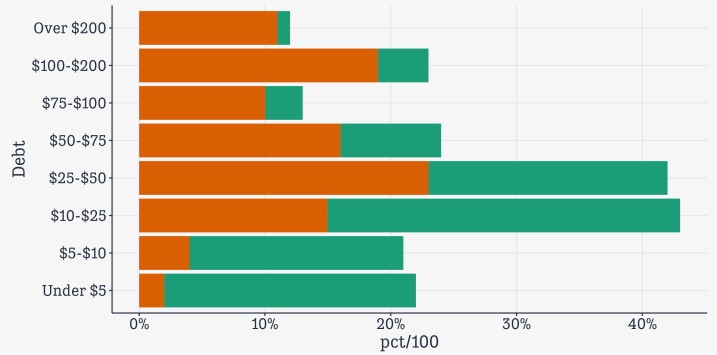
Debt Plot 1

```
studebt ▷
  ggplot(mapping =
           aes(x = pct/100,
               y = Debt,
               fill = type)) +
  geom_col() +
  scale_fill_brewer(type = "qual",
                    palette = "Dark2") +
  scale_x_continuous(labels = label_percent())
```



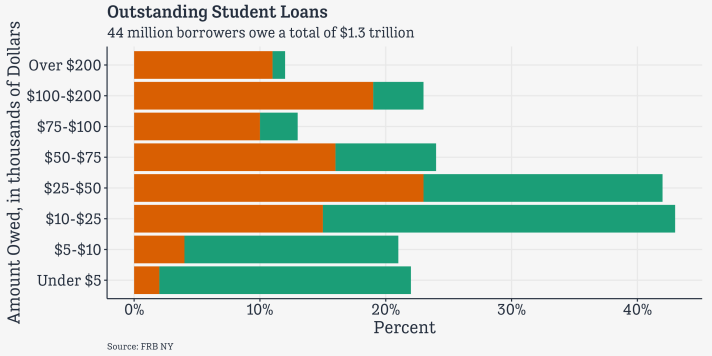
Debt Plot 1

```
studebt ▷
  ggplot(mapping =
           aes(x = pct/100,
               y = Debt,
               fill = type)) +
  geom_col() +
  scale_fill_brewer(type = "qual",
                    palette = "Dark2") +
  scale_x_continuous(labels = label_percent()) +
  guides(fill = "none")
```



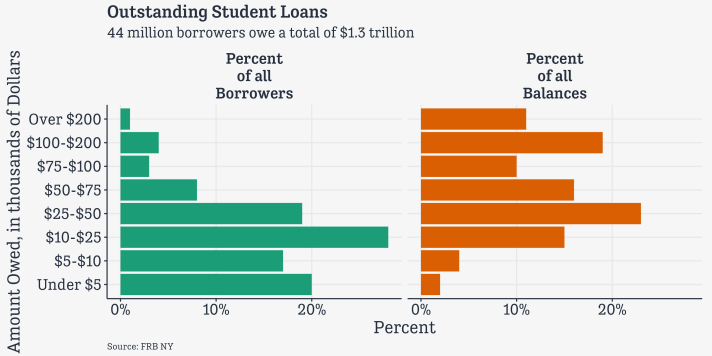
Debt Plot 1

```
studebt ▷
  ggplot(mapping =
           aes(x = pct/100,
               y = Debt,
               fill = type)) +
  geom_col() +
  scale_fill_brewer(type = "qual",
                    palette = "Dark2") +
  scale_x_continuous(labels = label_percent()) +
  guides(fill = "none") +
  labs(x = "Percent",
       y = p_ylab,
       caption = p_caption,
       title = p_title,
       subtitle = p_subtitle)
```



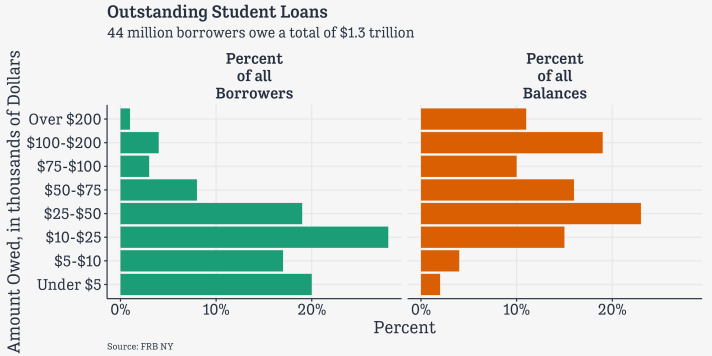
Debt Plot 1

```
studebt ▷
  ggplot(mapping =
           aes(x = pct/100,
               y = Debt,
               fill = type)) +
  geom_col() +
  scale_fill_brewer(type = "qual",
                    palette = "Dark2") +
  scale_x_continuous(labels = label_percent()) +
  guides(fill = "none") +
  labs(x = "Percent",
       y = p_ylab,
       caption = p_caption,
       title = p_title,
       subtitle = p_subtitle) +
  facet_wrap(~ type_label,
             labeller =
               label_wrap_gen(width=10))
```



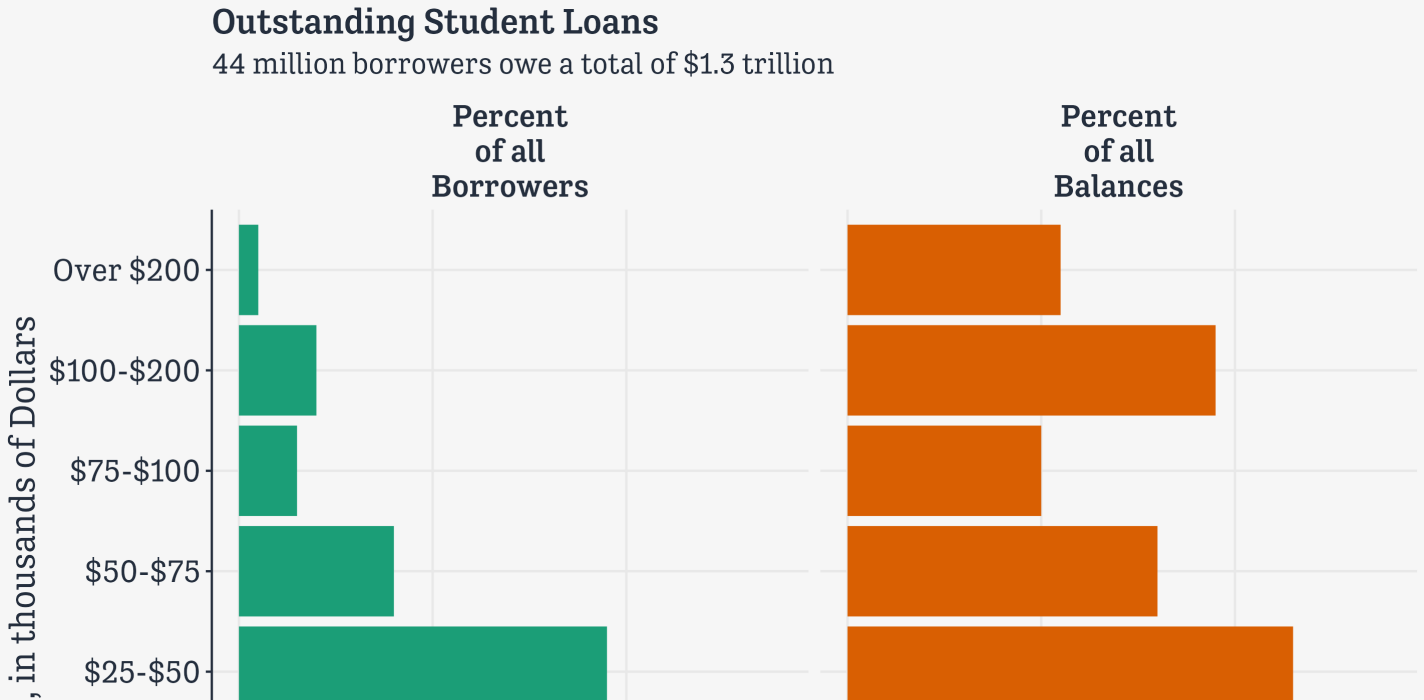
Debt Plot 1

```
studebt ▷
  ggplot(mapping =
           aes(x = pct/100,
               y = Debt,
               fill = type)) +
  geom_col() +
  scale_fill_brewer(type = "qual",
                    palette = "Dark2") +
  scale_x_continuous(labels = label_percent()) +
  guides(fill = "none") +
  labs(x = "Percent",
       y = p_ylab,
       caption = p_caption,
       title = p_title,
       subtitle = p_subtitle) +
  facet_wrap(~ type_label,
             labeller =
               label_wrap_gen(width=10)) +
    theme(strip.text.x =
          element_text(face = "bold"))
```



Debt Plot 1

```
studebt ▷
  ggplot(mapping =
           aes(x = pct/100,
               y = Debt,
               fill = type)) +
  geom_col() +
  scale_fill_brewer(type = "qual",
                    palette = "Dark2") +
  scale_x_continuous(labels = label_percent()) +
  guides(fill = "none") +
  labs(x = "Percent",
       y = p_ylab,
       caption = p_caption,
       title = p_title,
       subtitle = p_subtitle) +
  facet_wrap(~ type_label,
             labeller =
               label_wrap_gen(width=10)) +
    theme(strip.text.x =
          element_text(face = "bold")) →
  p1_debt
```


Outstanding Student Loans
44 million borrowers owe a total of $1.3 trillion
Percent of all Borrowers
Percent of all Balances
Over $200
$100-$200
$75-$100
$50-$75
$25-$50
in thousands of Dollars

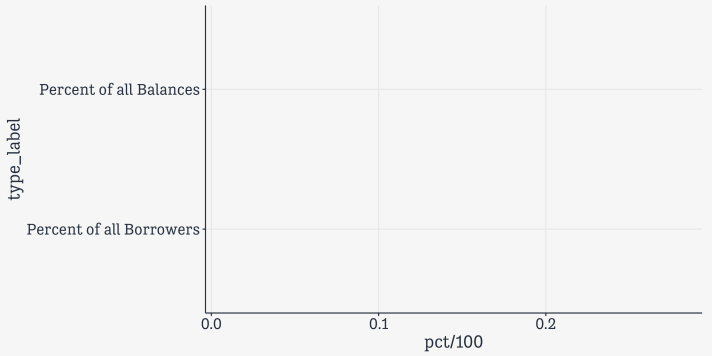
Alternatively, a kind of stacked bar chart

```
1 studebt
```

```
# A tibble: 16 × 5
   Debt      type        pct Debtrc    type_label
   <ord>     <fct>     <int> <ord>     <fct>
 1 Under $5  Borrowers    20 Under $5  Percent of
all Borrowers
 2 $5-$10    Borrowers    17 $5-$10    Percent of
all Borrowers
 3 $10-$25   Borrowers    28 $10-$25   Percent of
all Borrowers
 4 $25-$50   Borrowers    19 $25-$50   Percent of
all Borrowers
 5 $50-$75   Borrowers     8 $50-$75   Percent of
all Borrowers
 6 $75-$100  Borrowers     3 $75-$100  Percent of
all Borrowers
 7 $100-$200 Borrowers     4 $100-$200 Percent of
all Borrowers
 8 Over $200 Borrowers     1 Over $200 Percent of
all Borrowers
 9 Under $5  Balances      2 Under $5  Percent of
all Balances
10 $5-$10    Balances      4 $5-$10    Percent of
all Balances
11 $10-$25   Balances     15 $10-$25   Percent of
all Balances
```

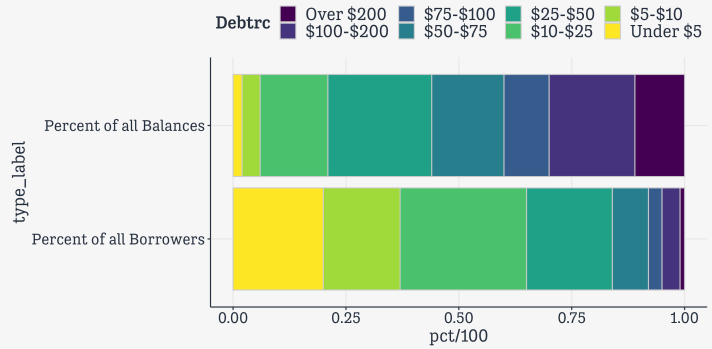
Alternatively, a kind of stacked bar chart

```
studebt ▷
  ggplot(mapping = aes(x = pct/100,
                       y = type_label,
                       fill = Debtrc))
```



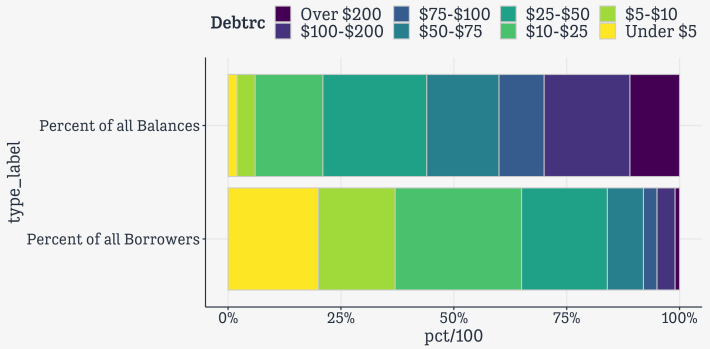
Alternatively, a kind of stacked bar chart

```
studebt ▷
  ggplot(mapping = aes(x = pct/100,
                       y = type_label,
                       fill = Debtrc)) +
  geom_col(color = "gray80")
```



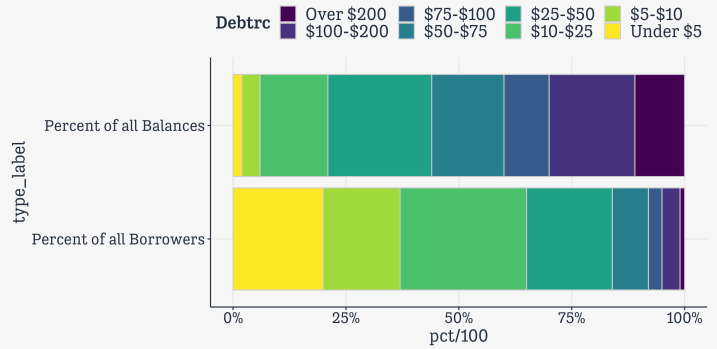
Alternatively, a kind of stacked bar chart

```
studebt ▷
  ggplot(mapping = aes(x = pct/100,
                       y = type_label,
                       fill = Debtrc)) +
  geom_col(color = "gray80") +
  scale_x_continuous(labels =
                       label_percent())
```



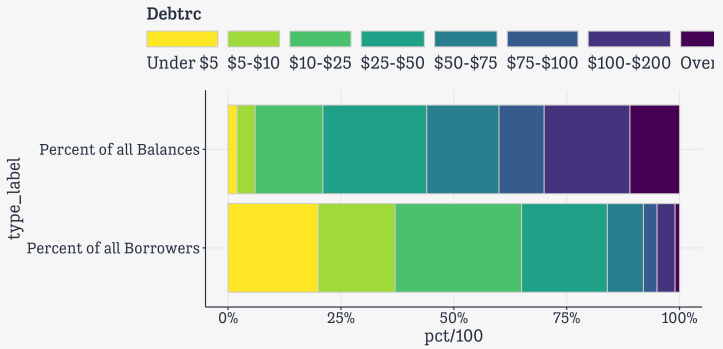
Alternatively, a kind of stacked bar chart

```
studebt ▷
  ggplot(mapping = aes(x = pct/100,
                       y = type_label,
                       fill = Debtrc)) +
  geom_col(color = "gray80") +
  scale_x_continuous(labels =
                       label_percent()) +
  scale_fill_viridis_d()
```



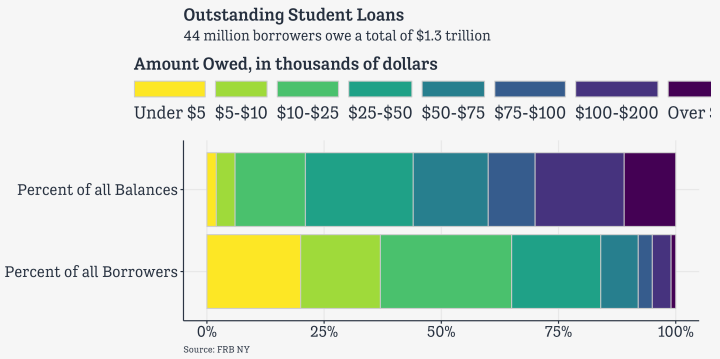
Alternatively, a kind of stacked bar chart

```
studebt ▷
  ggplot(mapping = aes(x = pct/100,
                       y = type_label,
                       fill = Debtrc)) +
  geom_col(color = "gray80") +
  scale_x_continuous(labels =
                       label_percent()) +
  scale_fill_viridis_d() +
  guides(fill =
           guide_legend(reverse = TRUE,
                        title.position = "top",
                        label.position = "bottom",
                        keywidth = 3,
                        nrow = 1))
```



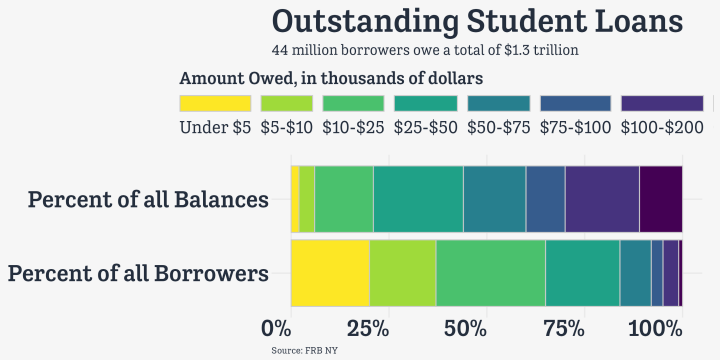
Alternatively, a kind of stacked bar chart

```
studebt ▷
  ggplot(mapping = aes(x = pct/100,
                       y = type_label,
                       fill = Debtrc)) +
  geom_col(color = "gray80") +
  scale_x_continuous(labels =
                       label_percent()) +
  scale_fill_viridis_d() +
  guides(fill =
           guide_legend(reverse = TRUE,
                        title.position = "top",
                        label.position = "bottom",
                        keywidth = 3,
                        nrow = 1)) +
  labs(x = NULL, y = NULL,
       fill = "Amount Owed, in thousands of dollars",
       caption = p_caption, title = p_title,
       subtitle = p_subtitle)
```



Alternatively, a kind of stacked bar chart

```
studebt ▷
  ggplot(mapping = aes(x = pct/100,
                       y = type_label,
                       fill = Debtrc)) +
  geom_col(color = "gray80") +
  scale_x_continuous(labels =
                       label_percent()) +
  scale_fill_viridis_d() +
  guides(fill =
           guide_legend(reverse = TRUE,
                        title.position = "top",
                        label.position = "bottom",
                        keywidth = 3,
                        nrow = 1)) +
  labs(x = NULL, y = NULL,
       fill = "Amount Owed, in thousands of dollars",
       caption = p_caption, title = p_title,
       subtitle = p_subtitle) +
  theme(legend.position = "top",
        plot.title = element_text(size = rel(2.8)),
        axis.text = element_text(face = "bold",
                hjust = 1,
                size = rel(2)),
```

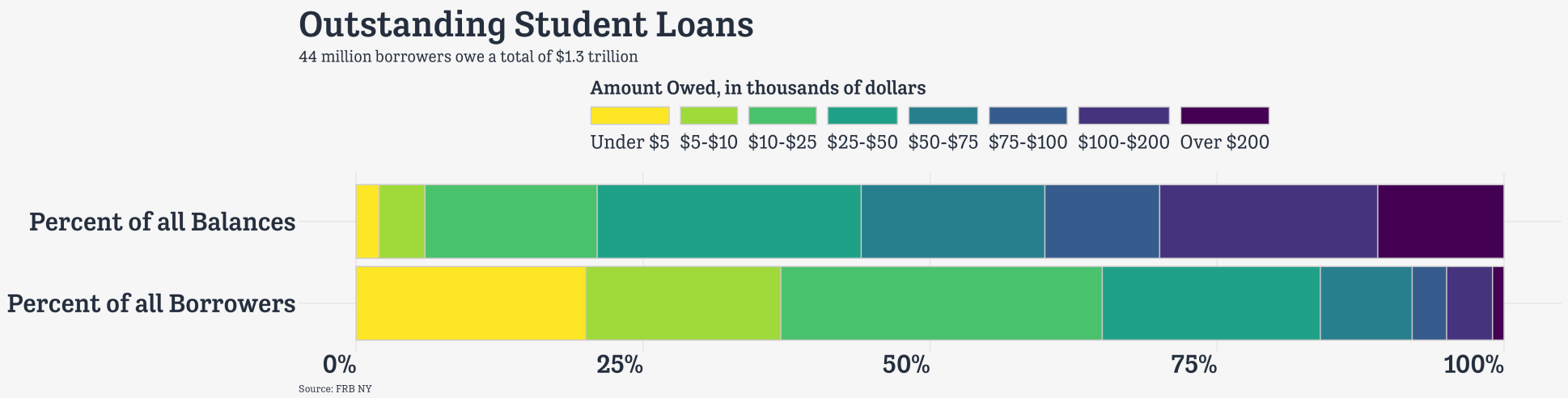


Alternatively, a kind of stacked bar chart

```
studebt ▷
  ggplot(mapping = aes(x = pct/100,
                       y = type_label,
                       fill = Debtrc)) +
  geom_col(color = "gray80") +
  scale_x_continuous(labels =
                       label_percent()) +
  scale_fill_viridis_d() +
  guides(fill =
           guide_legend(reverse = TRUE,
                        title.position = "top",
                        label.position = "bottom",
                        keywidth = 3,
                        nrow = 1)) +
  labs(x = NULL, y = NULL,
       fill = "Amount Owed, in thousands of dollars",
       caption = p_caption, title = p_title,
       subtitle = p_subtitle) +
  theme(legend.position = "top",
        plot.title = element_text(size = rel(2.8)),
        axis.text = element_text(face = "bold",
                hjust = 1,
                size = rel(2)),
```

Alternatively, a kind of stacked bar chart

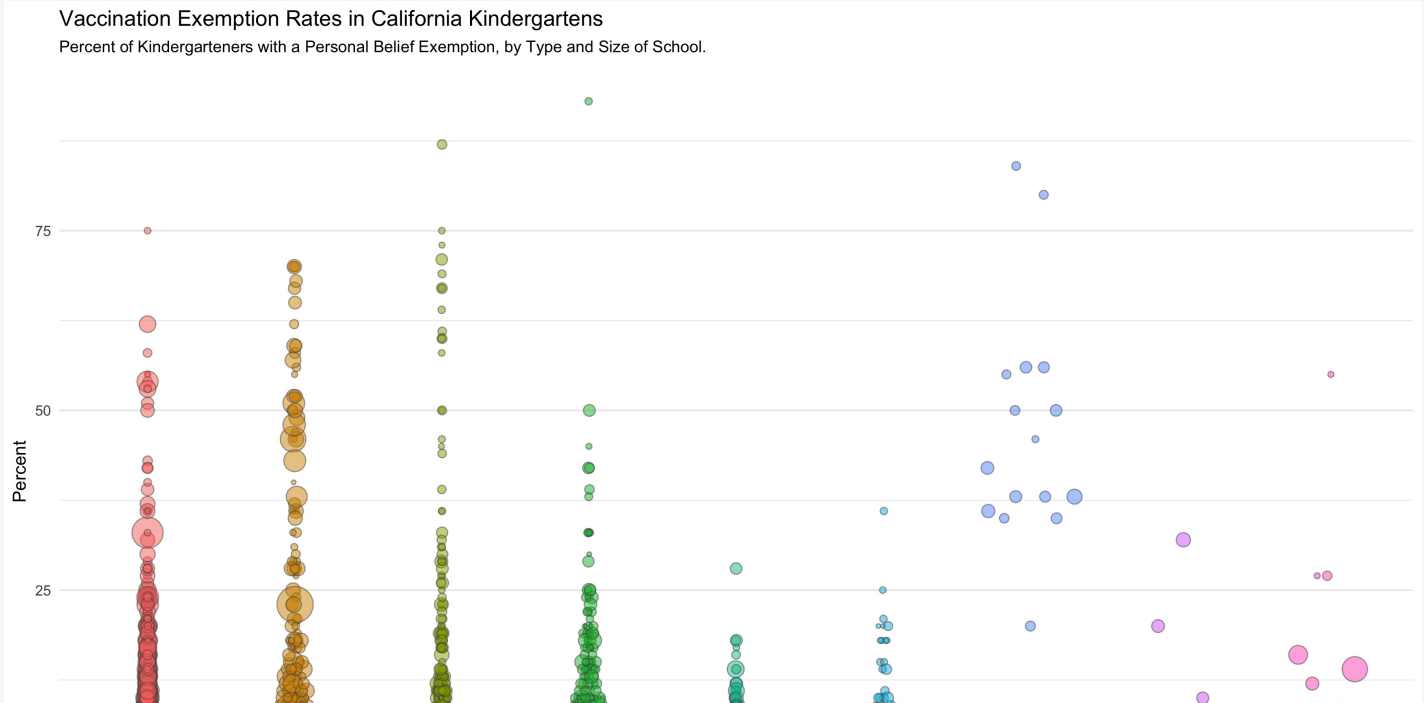
```
studebt ▷
  ggplot(mapping = aes(x = pct/100,
                       y = type_label,
                       fill = Debtrc)) +
  geom_col(color = "gray80") +
  scale_x_continuous(labels =
                       label_percent()) +
  scale_fill_viridis_d() +
  guides(fill =
           guide_legend(reverse = TRUE,
                        title.position = "top",
                        label.position = "bottom",
                        keywidth = 3,
                        nrow = 1)) +
  labs(x = NULL, y = NULL,
       fill = "Amount Owed, in thousands of dollars",
       caption = p_caption, title = p_title,
       subtitle = p_subtitle) +
  theme(legend.position = "top",
        plot.title = element_text(size = rel(2.8)),
        axis.text = element_text(face = "bold",
                hjust = 1,
                size = rel(2)),
```



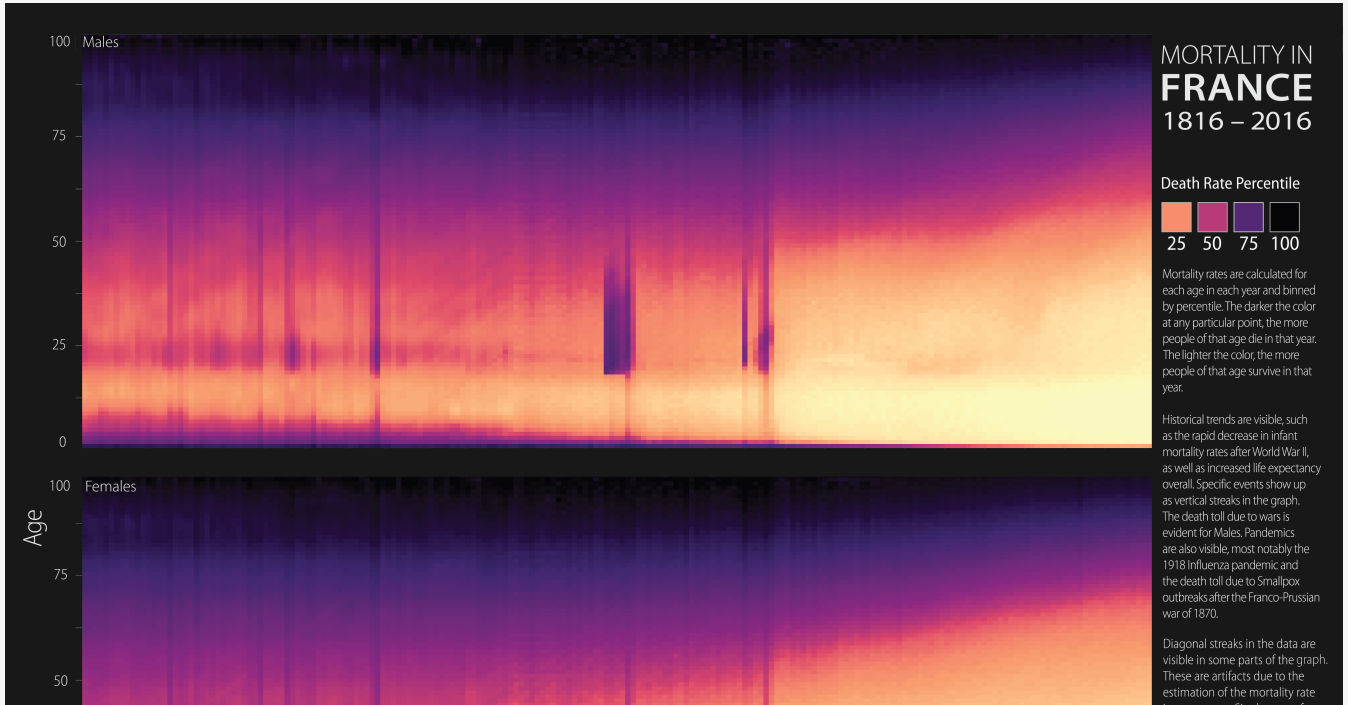
Pies redrawn as sideways-stacked columns

But I want a pony

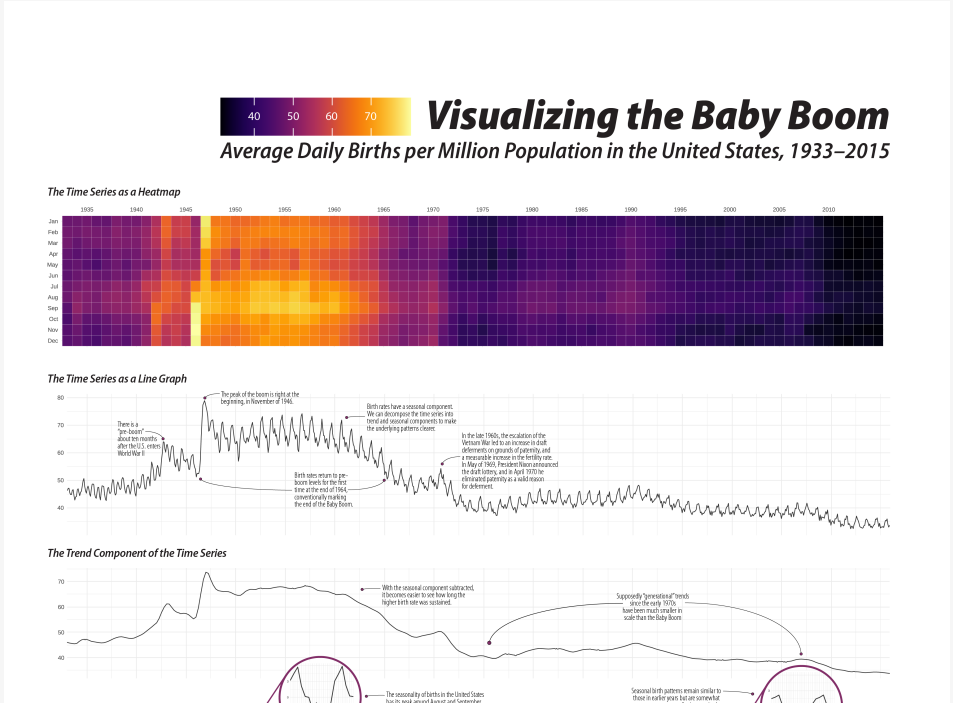
Show ponies


Vaccination Exemption Rates in California Kindergartens
Percent of Kindergarteners with a Personal Belief Exemption, by Type and Size of School.
Percent
75
50
25

Show ponies


100
Males
75
50
25
0
100
Females
Age
75
50
MORTALITY IN
FRANCE
1816 – 2016
Death Rate Percentile
25 50 75 100
Mortality rates are calculated for each age in each year and binned by percentile. The darker the color at any particular point, the more people of that age die in that year. The lighter the color, the more people of that age survive in that year.
Historical trends are visible, such as the rapid decrease in infant mortality rates after World War II, as well as increased life expectancy overall. Specific events show up as vertical streaks in the graph. The death toll due to wars is evident for Males. Pandemics are also visible, most notably the 1918 Influenza pandemic and the death toll due to Smallpox outbreaks after the Franco-Prussian war of 1870.
Diagonal streaks in the data are visible in some parts of the graph. These are artifacts due to the estimation of the mortality rate

Show ponies


40
50
60
70
Visualizing the Baby Boom
Average Daily Births per Million Population in the United States, 1933–2015
The Time Series as a Heatmap
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
The Time Series as a Line Graph
80
70
60
50
40
There is a "pre-boom" — about ten months after the U.S. enters World War II
The peak of the boom is right at the beginning, in November of 1946.
Birth rates have a seasonal component. We can decompose the time series into trend and seasonal components to make the underlying patterns clearer.
Birth rates return to pre-boom levels for the first time at the end of 1964, conventionally marking the end of the Baby Boom.
In the late 1960s, the escalation of the Vietnam War led to an increase in draft deferments on grounds of paternity, and a measurable increase in the fertility rate. In May of 1969, President Nixon announced the draft lottery, and in April 1970 he eliminated paternity as a valid reason for deferment.
The Trend Component of the Time Series
70
60
50
40
With the seasonal component subtracted, it becomes easier to see how long the higher birth rate was sustained.
Supposedly "generational" trends since the early 1970s have been much smaller in scale than the Baby Boom.
The seasonality of births in the United States
Seasonal birth patterns remain similar to those in earlier years but are somewhat

OK boomer

The demog package

```
# remotes::install_github("kjhealy/demog")
# library(demog)

okboomer
```

```
# A tibble: 1,644 × 12
    year month n_days births total_pop births_pct births_pct_day date
   <dbl> <dbl>  <dbl>  <dbl>     <dbl>      <dbl>          <dbl> <date>
 1  1938     1     31  51820  41215000    0.00126           40.6 1938-01-01
 2  1938     2     28  47421  41215000    0.00115           41.1 1938-02-01
 3  1938     3     31  54887  41215000    0.00133           43.0 1938-03-01
 4  1938     4     30  54623  41215000    0.00133           44.2 1938-04-01
 5  1938     5     31  56853  41215000    0.00138           44.5 1938-05-01
 6  1938     6     30  53145  41215000    0.00129           43.0 1938-06-01
 7  1938     7     31  53214  41215000    0.00129           41.6 1938-07-01
 8  1938     8     31  50444  41215000    0.00122           39.5 1938-08-01
 9  1938     9     30  50545  41215000    0.00123           40.9 1938-09-01
10  1938    10     31  50079  41215000    0.00122           39.2 1938-10-01
# ℹ 1,634 more rows
# ℹ 4 more variables: seasonal <dbl>, trend <dbl>, remainder <dbl>,
#   country <chr>
```

Boomer Line Graph

```
1  okboomer
```

```
# A tibble: 1,644 × 12
    year month n_days births total_pop births_pct
births_pct_day date
   <dbl> <dbl>  <dbl>  <dbl>     <dbl>      <dbl>
<dbl> <date>
 1  1938     1     31  51820  41215000    0.00126
40.6 1938-01-01
 2  1938     2     28  47421  41215000    0.00115
41.1 1938-02-01
 3  1938     3     31  54887  41215000    0.00133
43.0 1938-03-01
 4  1938     4     30  54623  41215000    0.00133
44.2 1938-04-01
 5  1938     5     31  56853  41215000    0.00138
44.5 1938-05-01
 6  1938     6     30  53145  41215000    0.00129
43.0 1938-06-01
 7  1938     7     31  53214  41215000    0.00129
41.6 1938-07-01
 8  1938     8     31  50444  41215000    0.00122
39.5 1938-08-01
 9  1938     9     30  50545  41215000    0.00123
40.9 1938-09-01
10  1938    10     31  50079  41215000    0.00122
39.2 1938-10-01
```

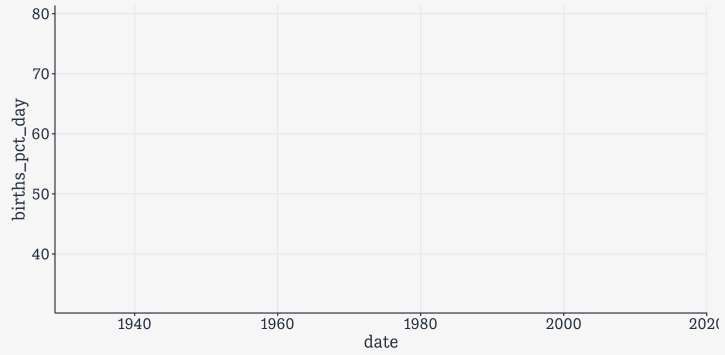
Boomer Line Graph

```
okboomer ▷
    filter(country == "United States")
```

```
# A tibble: 996 × 12
    year month n_days births total_pop births_pct
births_pct_day date
   <dbl> <dbl>  <dbl>  <dbl>     <dbl>      <dbl>
<dbl> <date>
 1  1933     1     31 180545 125579000    0.00144
46.4 1933-01-01
 2  1933     2     28 165986 125579000    0.00132
47.2 1933-02-01
 3  1933     3     31 183762 125579000    0.00146
47.2 1933-03-01
 4  1933     4     30 171354 125579000    0.00136
45.5 1933-04-01
 5  1933     5     31 174811 125579000    0.00139
44.9 1933-05-01
 6  1933     6     30 169255 125579000    0.00135
44.9 1933-06-01
 7  1933     7     31 180880 125579000    0.00144
46.5 1933-07-01
 8  1933     8     31 181856 125579000    0.00145
46.7 1933-08-01
 9  1933     9     30 167637 125579000    0.00133
44.5 1933-09-01
10  1933    10     31 167055 125579000    0.00133
42.9 1933-10-01
```

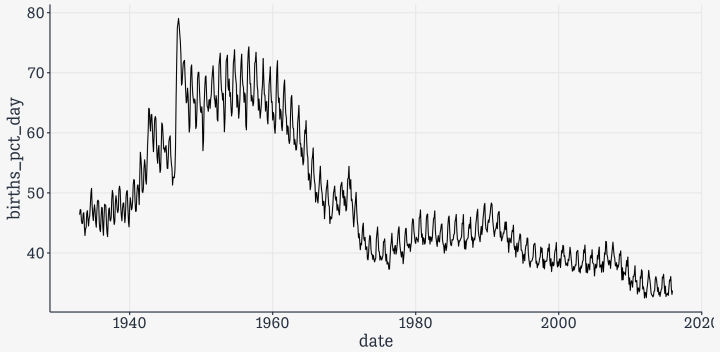
Boomer Line Graph

```
okboomer |>
    filter(country == "United States") |>
    ggplot(aes(x = date, y = births_pct_day))
```



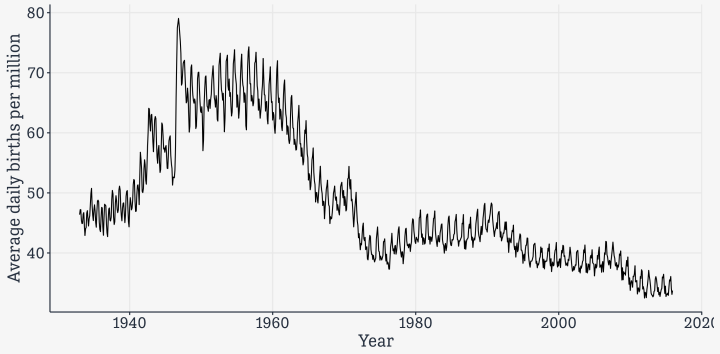
Boomer Line Graph

```
okboomer |>
    filter(country == "United States")  |>
    ggplot(aes(x = date, y = births_pct_day)) +
    geom_line(linewidth = 0.5)
```



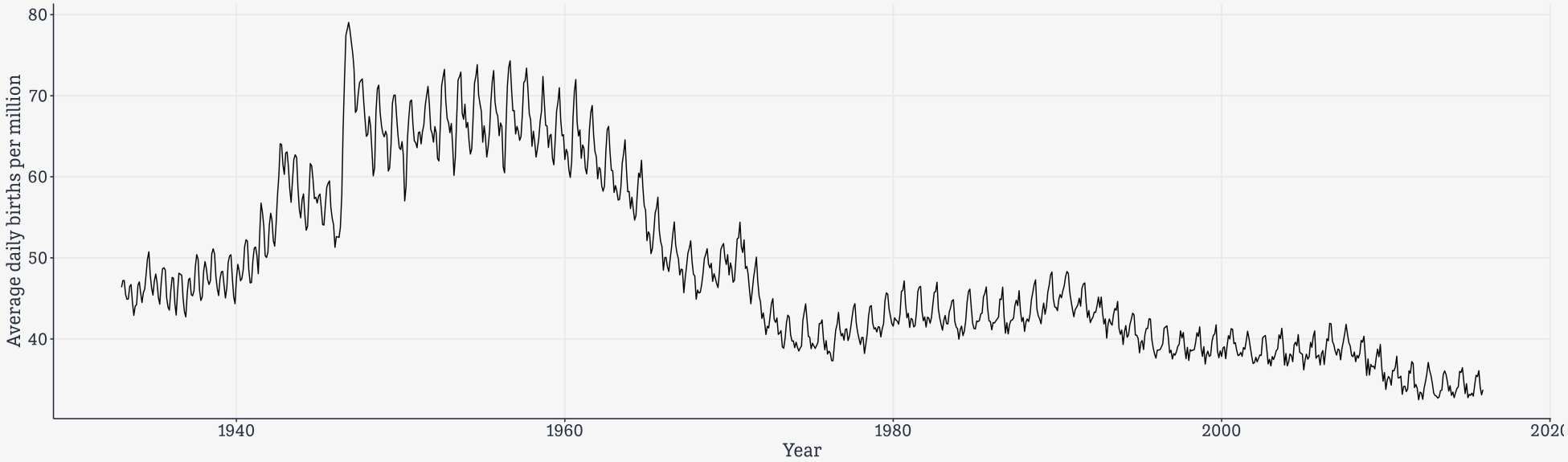
Boomer Line Graph

```
okboomer ▷
    filter(country == "United States")  ▷
    ggplot(aes(x = date, y = births_pct_day)) +
    geom_line(linewidth = 0.5) +
    labs(x = "Year",
         y = "Average daily births per million")
```



Boomer Line Graph

```
okboomer |>
    filter(country == "United States")  |>
    ggplot(aes(x = date, y = births_pct_day)) +
    geom_line(linewidth = 0.5) +
    labs(x = "Year",
         y = "Average daily births per million") ->
  p_lineboom
```



The Baby Boom.

Tiled Heatmap

```
1  okboomer
```

```
# A tibble: 1,644 × 12
    year month n_days births total_pop births_pct
births_pct_day date
   <dbl> <dbl>  <dbl>  <dbl>     <dbl>      <dbl>
<dbl> <date>
 1  1938     1     31  51820  41215000    0.00126
40.6 1938-01-01
 2  1938     2     28  47421  41215000    0.00115
41.1 1938-02-01
 3  1938     3     31  54887  41215000    0.00133
43.0 1938-03-01
 4  1938     4     30  54623  41215000    0.00133
44.2 1938-04-01
 5  1938     5     31  56853  41215000    0.00138
44.5 1938-05-01
 6  1938     6     30  53145  41215000    0.00129
43.0 1938-06-01
 7  1938     7     31  53214  41215000    0.00129
41.6 1938-07-01
 8  1938     8     31  50444  41215000    0.00122
39.5 1938-08-01
 9  1938     9     30  50545  41215000    0.00123
40.9 1938-09-01
10  1938    10     31  50079  41215000    0.00122
39.2 1938-10-01
```

Tiled Heatmap

```
okboomer ▷
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE))
```

```
# A tibble: 1,644 × 14
    year month n_days births total_pop births_pct
births_pct_day date
   <dbl> <dbl>  <dbl>  <dbl>     <dbl>      <dbl>
<dbl> <date>
 1  1938     1     31  51820  41215000    0.00126
40.6 1938-01-01
 2  1938     2     28  47421  41215000    0.00115
41.1 1938-02-01
 3  1938     3     31  54887  41215000    0.00133
43.0 1938-03-01
 4  1938     4     30  54623  41215000    0.00133
44.2 1938-04-01
 5  1938     5     31  56853  41215000    0.00138
44.5 1938-05-01
 6  1938     6     30  53145  41215000    0.00129
43.0 1938-06-01
 7  1938     7     31  53214  41215000    0.00129
41.6 1938-07-01
 8  1938     8     31  50444  41215000    0.00122
39.5 1938-08-01
 9  1938     9     30  50545  41215000    0.00123
40.9 1938-09-01
10  1938    10     31  50079  41215000    0.00122
39.2 1938-10-01
```

Tiled Heatmap

```
okboomer ▷
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) ▷
    select(year, month, year_fct, month_fct, everything())
```

```
# A tibble: 1,644 × 14
    year month year_fct month_fct n_days births
total_pop births_pct
   <dbl> <dbl> <ord>    <ord>      <dbl>  <dbl>
<dbl>      <dbl>
 1  1938     1 1938     Jan           31  51820
41215000    0.00126
 2  1938     2 1938     Feb           28  47421
41215000    0.00115
 3  1938     3 1938     Mar           31  54887
41215000    0.00133
 4  1938     4 1938     Apr           30  54623
41215000    0.00133
 5  1938     5 1938     May           31  56853
41215000    0.00138
 6  1938     6 1938     Jun           30  53145
41215000    0.00129
 7  1938     7 1938     Jul           31  53214
41215000    0.00129
 8  1938     8 1938     Aug           31  50444
41215000    0.00122
 9  1938     9 1938     Sep           30  50545
41215000    0.00123
10  1938    10 1938     Oct           31  50079
41215000    0.00122
```

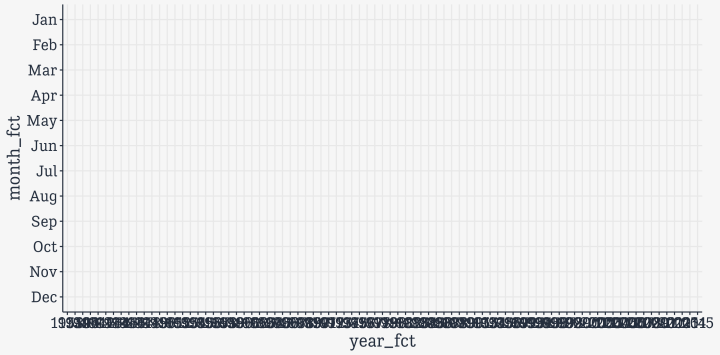
Tiled Heatmap

```
okboomer ▷
    mutate(year_fct =
               factor(year,
                      levels = unique(year),
                      ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) ▷
    select(year, month, year_fct, month_fct, everything()) ▷
  filter(country == "United States")
```

```
# A tibble: 996 × 14
    year month year_fct month_fct n_days births
total_pop births_pct
   <dbl> <dbl> <ord>    <ord>      <dbl>  <dbl>
<dbl>      <dbl>
 1  1933     1 1933     Jan           31 180545
125579000    0.00144
 2  1933     2 1933     Feb           28 165986
125579000    0.00132
 3  1933     3 1933     Mar           31 183762
125579000    0.00146
 4  1933     4 1933     Apr           30 171354
125579000    0.00136
 5  1933     5 1933     May           31 174811
125579000    0.00139
 6  1933     6 1933     Jun           30 169255
125579000    0.00135
 7  1933     7 1933     Jul           31 180880
125579000    0.00144
 8  1933     8 1933     Aug           31 181856
125579000    0.00145
 9  1933     9 1933     Sep           30 167637
125579000    0.00133
10  1933    10 1933     Oct           31 167055
125579000    0.00133
```

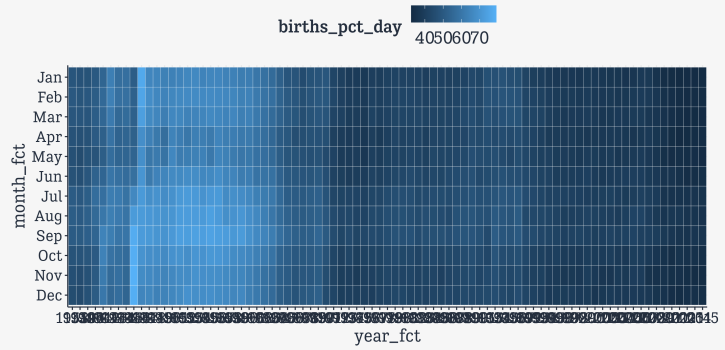
Tiled Heatmap

```r
okboomer |>
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) |>
    select(year, month, year_fct, month_fct, everything()) |>
  filter(country == "United States") |>
    ggplot(aes(x = year_fct, y = month_fct))
```



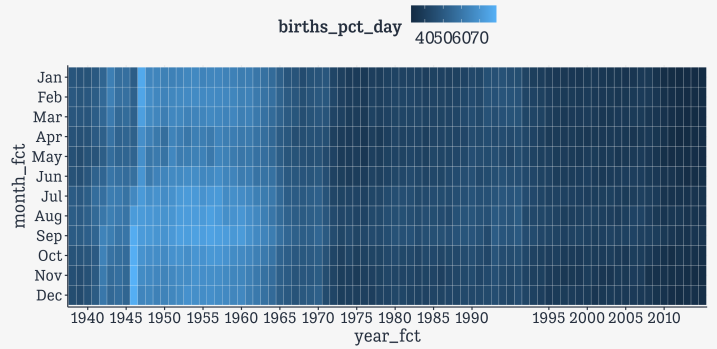
Tiled Heatmap

```
okboomer ▷
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) ▷
    select(year, month, year_fct, month_fct, everything()) ▷
  filter(country == "United States") ▷
    ggplot(aes(x = year_fct, y = month_fct)) +
    geom_tile(mapping = aes(fill = births_pct_day),
              color = "white")
```



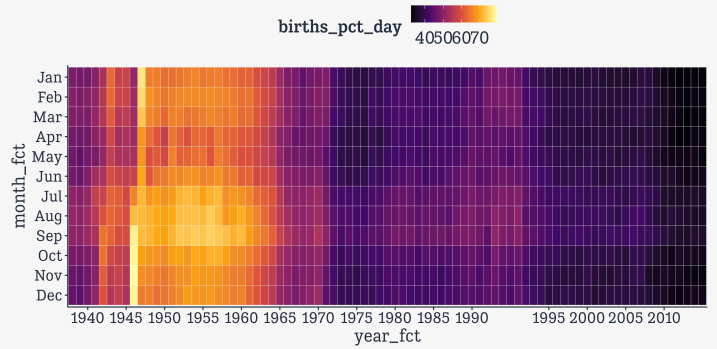
Tiled Heatmap

```
okboomer ▷
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) ▷
    select(year, month, year_fct, month_fct, everything()) ▷
  filter(country == "United States") ▷
    ggplot(aes(x = year_fct, y = month_fct)) +
    geom_tile(mapping = aes(fill = births_pct_day),
              color = "white") +
   scale_x_discrete(breaks = seq(1940, 2010, 5))
```



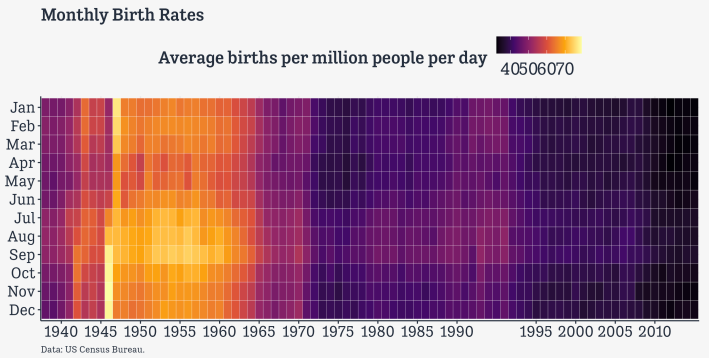
Tiled Heatmap

```
okboomer ▷
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) ▷
    select(year, month, year_fct, month_fct, everything()) ▷
  filter(country == "United States") ▷
    ggplot(aes(x = year_fct, y = month_fct)) +
    geom_tile(mapping = aes(fill = births_pct_day),
              color = "white") +
   scale_x_discrete(breaks = seq(1940, 2010, 5)) +
   scale_fill_viridis_c(option = "B")
```



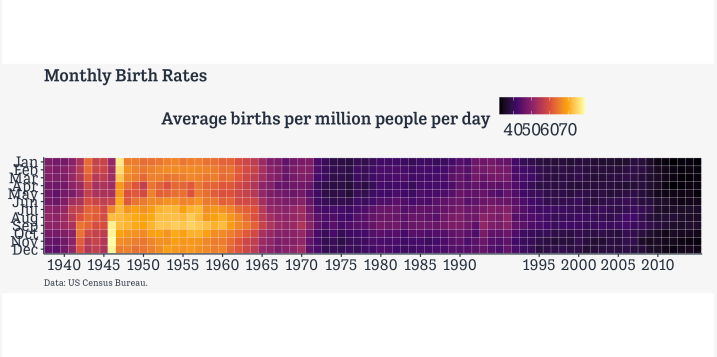
Tiled Heatmap

```
okboomer ▷
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) ▷
    select(year, month, year_fct, month_fct, everything()) ▷
  filter(country == "United States") ▷
    ggplot(aes(x = year_fct, y = month_fct)) +
    geom_tile(mapping = aes(fill = births_pct_day),
              color = "white") +
   scale_x_discrete(breaks = seq(1940, 2010, 5)) +
   scale_fill_viridis_c(option = "B") +
  labs(x = NULL, y = NULL,
       title = "Monthly Birth Rates",
       fill = "Average births per million people per day",
         caption = "Data: US Census Bureau.")
```



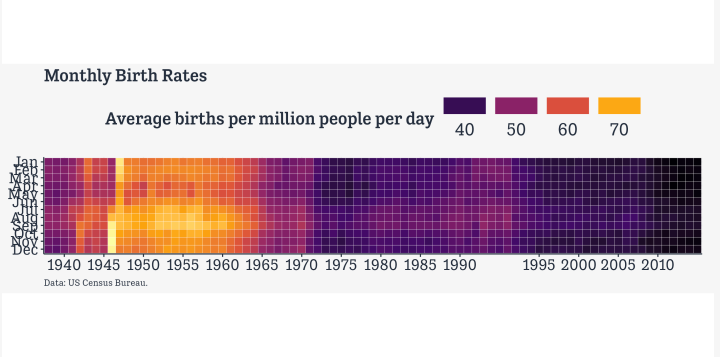
Tiled Heatmap

```r
okboomer |>
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) |>
    select(year, month, year_fct, month_fct, everything()) |>
  filter(country == "United States") |>
    ggplot(aes(x = year_fct, y = month_fct)) +
    geom_tile(mapping = aes(fill = births_pct_day),
              color = "white") +
   scale_x_discrete(breaks = seq(1940, 2010, 5)) +
   scale_fill_viridis_c(option = "B") +
  labs(x = NULL, y = NULL,
       title = "Monthly Birth Rates",
       fill = "Average births per million people per day",
         caption = "Data: US Census Bureau.") +
  coord_fixed()
```



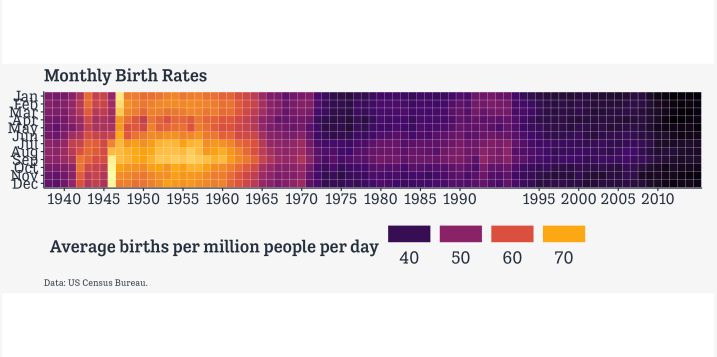
Tiled Heatmap

```
okboomer ▷
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) ▷
    select(year, month, year_fct, month_fct, everything()) ▷
  filter(country == "United States") ▷
    ggplot(aes(x = year_fct, y = month_fct)) +
    geom_tile(mapping = aes(fill = births_pct_day),
              color = "white") +
   scale_x_discrete(breaks = seq(1940, 2010, 5)) +
   scale_fill_viridis_c(option = "B") +
  labs(x = NULL, y = NULL,
       title = "Monthly Birth Rates",
       fill = "Average births per million people per day",
         caption = "Data: US Census Bureau.") +
  coord_fixed() +
```



Tiled Heatmap

```
okboomer ▷
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) ▷
    select(year, month, year_fct, month_fct, everything()) ▷
  filter(country == "United States") ▷
    ggplot(aes(x = year_fct, y = month_fct)) +
    geom_tile(mapping = aes(fill = births_pct_day),
              color = "white") +
   scale_x_discrete(breaks = seq(1940, 2010, 5)) +
   scale_fill_viridis_c(option = "B") +
  labs(x = NULL, y = NULL,
       title = "Monthly Birth Rates",
       fill = "Average births per million people per day",
         caption = "Data: US Census Bureau.") +
  coord_fixed() +
```

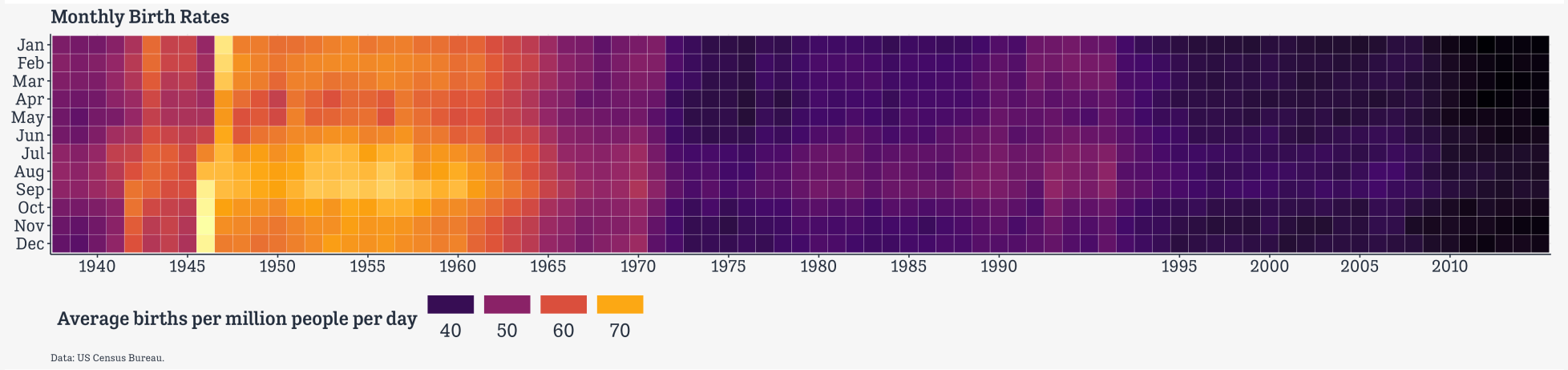


Tiled Heatmap

```
okboomer ▷
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) ▷
    select(year, month, year_fct, month_fct, everything()) ▷
  filter(country == "United States") ▷
    ggplot(aes(x = year_fct, y = month_fct)) +
    geom_tile(mapping = aes(fill = births_pct_day),
              color = "white") +
   scale_x_discrete(breaks = seq(1940, 2010, 5)) +
   scale_fill_viridis_c(option = "B") +
  labs(x = NULL, y = NULL,
       title = "Monthly Birth Rates",
       fill = "Average births per million people per day",
         caption = "Data: US Census Bureau.") +
  coord_fixed() +
```

Tiled Heatmap

```
okboomer |>
    mutate(year_fct =
             factor(year,
                    levels = unique(year),
                    ordered = TRUE),
           month_fct = factor(month,
                              levels = rev(c(1:12)),
                              labels = rev(c("Jan", "Feb", "Mar", "Apr",
                                  "May", "Jun", "Jul", "Aug",
                                  "Sep", "Oct", "Nov", "Dec")),
                              ordered = TRUE)) |>
    select(year, month, year_fct, month_fct, everything()) |>
  filter(country == "United States") |>
    ggplot(aes(x = year_fct, y = month_fct)) +
    geom_tile(mapping = aes(fill = births_pct_day),
              color = "white") +
   scale_x_discrete(breaks = seq(1940, 2010, 5)) +
   scale_fill_viridis_c(option = "B") +
  labs(x = NULL, y = NULL,
       title = "Monthly Birth Rates",
       fill = "Average births per million people per day",
         caption = "Data: US Census Bureau.") +
  coord_fixed() +
```



The Baby Boom as a tiled temporal heatmap

Beeswarms and bespoke labels

The cavax package

```
# remotes::install_github("kjhealy/cavax)
library(cavax)

cavax
```

```
# A tibble: 7,032 × 13
     code county name  type  district city  enrollment pbe_pct exempt med_exempt
    <dbl> <chr>  <chr> <chr> <chr>    <chr>       <dbl>   <dbl>  <dbl>      <dbl>
 1 1.10e5 ALAME… FAME… PUBL… ALAMEDA… NEWA…        109      13  12.8        0
 2 6.00e6 ALAME… COX … PUBL… ALAMEDA… OAKL…        115       1   0.87       0.87
 3 6.00e6 ALAME… LAZE… PUBL… ALAMEDA… OAKL…         40       0   0          0
 4 1.24e5 ALAME… YU M… PUBL… ALAMEDA… OAKL…         52      10   9.62       0
 5 6.10e6 ALAME… AMEL… PUBL… ALAMEDA… ALAM…        128       2   1.56       0
 6 6.11e6 ALAME… BAY … PUBL… ALAMEDA… ALAM…         70       1   1.43       0
 7 6.09e6 ALAME… DONA… PUBL… ALAMEDA… ALAM…        100       3   3          0
 8 6.09e6 ALAME… EDIS… PUBL… ALAMEDA… ALAM…         70       1   1.43       0
 9 6.09e6 ALAME… FRAN… PUBL… ALAMEDA… ALAM…         95       1   1.05       1.05
10 6.09e6 ALAME… FRAN… PUBL… ALAMEDA… ALAM…         50       2   2          0
# ℹ 7,022 more rows
# ℹ 3 more variables: rel_exempt <dbl>, mwc <fct>, kind <fct>
```

Here we will do some custom manual labeling.

Aux Info Panel

```
1  library(ggbeeswarm)
```

Aux Info Panel

```
1  library(ggbeeswarm)
2  make_comma ← scales::label_comma()
```

Aux Info Panel

```
1 library(ggbeeswarm)
2 make_comma ← scales::label_comma()
3
4 cavax
```

```
# A tibble: 7,032 × 13
     code county name  type  district city
enrollment pbe_pct exempt med_exempt
    <dbl> <chr>  <chr> <chr> <chr>    <chr>
<dbl>   <dbl>  <dbl>      <dbl>
 1 1.10e5 ALAME… FAME… PUBL… ALAMEDA… NEWA…
109      13  12.8         0
 2 6.00e6 ALAME… COX … PUBL… ALAMEDA… OAKL…
115       1   0.87        0.87
 3 6.00e6 ALAME… LAZE… PUBL… ALAMEDA… OAKL…
40        0   0           0
 4 1.24e5 ALAME… YU M… PUBL… ALAMEDA… OAKL…
52       10   9.62        0
 5 6.10e6 ALAME… AMEL… PUBL… ALAMEDA… ALAM…
128       2   1.56        0
 6 6.11e6 ALAME… BAY … PUBL… ALAMEDA… ALAM…
70        1   1.43        0
 7 6.09e6 ALAME… DONA… PUBL… ALAMEDA… ALAM…
100       3   3           0
 8 6.09e6 ALAME… EDIS… PUBL… ALAMEDA… ALAM…
70        1   1.43        0
 9 6.09e6 ALAME… FRAN… PUBL… ALAMEDA… ALAM…
95        1   1.05        1.05
10 6.09e6 ALAME… FRAN… PUBL… ALAMEDA… ALAM…
50        2   2           0
```

Aux Info Panel

```
1 library(ggbeeswarm)
2 make_comma ← scales::label_comma()
3
4 cavax ▷
5   group_by(mwc)
```

```
# A tibble: 7,032 × 13
# Groups:   mwc [11]
     code county name  type  district city
enrollment pbe_pct exempt med_exempt
    <dbl> <chr>  <chr> <chr> <chr>    <chr>
<dbl>   <dbl>  <dbl>      <dbl>
 1 1.10e5 ALAME… FAME… PUBL… ALAMEDA… NEWA…
109      13  12.8        0
 2 6.00e6 ALAME… COX … PUBL… ALAMEDA… OAKL…
115       1   0.87       0.87
 3 6.00e6 ALAME… LAZE… PUBL… ALAMEDA… OAKL…
40        0   0          0
 4 1.24e5 ALAME… YU M… PUBL… ALAMEDA… OAKL…
52       10   9.62       0
 5 6.10e6 ALAME… AMEL… PUBL… ALAMEDA… ALAM…
128       2   1.56       0
 6 6.11e6 ALAME… BAY … PUBL… ALAMEDA… ALAM…
70        1   1.43       0
 7 6.09e6 ALAME… DONA… PUBL… ALAMEDA… ALAM…
100       3   3          0
 8 6.09e6 ALAME… EDIS… PUBL… ALAMEDA… ALAM…
70        1   1.43       0
 9 6.09e6 ALAME… FRAN… PUBL… ALAMEDA… ALAM…
95        1   1.05       1.05
10 6.09e6 ALAME… FRAN… PUBL… ALAMEDA… ALAM…
```

Aux Info Panel

```
1  library(ggbeeswarm)
2  make_comma ← scales::label_comma()
3
4  cavax ▷
5    group_by(mwc) ▷
6    summarize(n_schools=n(),
7              n_students = sum(enrollment, na.rm=TRUE))
```

```
# A tibble: 11 × 3
   mwc                          n_schools n_students
   <fct>                            <int>      <dbl>
 1 Public                            5314     472802
 2 Charter                            314      19863
 3 Private Non-Specific               591      16697
 4 Private Christian                  336       8836
 5 Private Catholic                   334       9869
 6 Private Montessori                  99       2112
 7 Private Waldorf                     16        513
 8 Charter Montessori                   5        227
 9 Public Montessori                   11        706
10 Private Christian Montessori         4         78
11 Private Jewish/Islamic               8        237
```

Aux Info Panel

```
1  library(ggbeeswarm)
2  make_comma ← scales::label_comma()
3
4  cavax ▷
5    group_by(mwc) ▷
6    summarize(n_schools=n(),
7              n_students = sum(enrollment, na.rm=TRUE)) ▷
8    drop_na()
```

```
# A tibble: 11 × 3
   mwc                          n_schools n_students
   <fct>                            <int>      <dbl>
 1 Public                            5314     472802
 2 Charter                            314      19863
 3 Private Non-Specific               591      16697
 4 Private Christian                  336       8836
 5 Private Catholic                   334       9869
 6 Private Montessori                  99       2112
 7 Private Waldorf                     16        513
 8 Charter Montessori                   5        227
 9 Public Montessori                   11        706
10 Private Christian Montessori         4         78
11 Private Jewish/Islamic               8        237
```

Aux Info Panel

```r
library(ggbeeswarm)
make_comma <- scales::label_comma()

cavax |>
  group_by(mwc) |>
  summarize(n_schools=n(),
            n_students = sum(enrollment, na.rm=TRUE)) |>
  drop_na() |>
  mutate(n_schools_fmt = make_comma(n_schools),
         n_students_fmt = make_comma(n_students),
         info_schools = paste(n_schools_fmt, "Schools Enrolling"),
         info_students = paste(n_students_fmt, "Kindergarteners"))
```

```
# A tibble: 11 × 7
   mwc            n_schools n_students n_schools_fmt
n_students_fmt info_schools
   <fct>              <int>      <dbl> <chr>
<chr>          <chr>
 1 Public              5314     472802 5,314
472,802        5,314 Schoo…
 2 Charter              314      19863 314
19,863         314 Schools…
 3 Private Non-S…       591      16697 591
16,697         591 Schools…
 4 Private Chris…       336       8836 336
8,836          336 Schools…
 5 Private Catho…       334       9869 334
9,869          334 Schools…
 6 Private Monte…        99       2112 99
2,112          99 Schools …
 7 Private Waldo…        16        513 16
513            16 Schools …
 8 Charter Monte…         5        227 5
227            5 Schools E…
 9 Public Montes…        11        706 11
706            11 Schools …
10 Private Chris…         4         78 4
78             4 Schools E…
```

Aux Info Panel

```
library(ggbeeswarm)
make_comma ← scales::label_comma()

cavax ▷
  group_by(mwc) ▷
  summarize(n_schools=n(),
            n_students = sum(enrollment, na.rm=TRUE)) ▷
  drop_na() ▷
  mutate(n_schools_fmt = make_comma(n_schools),
         n_students_fmt = make_comma(n_students),
         info_schools = paste(n_schools_fmt, "Schools Enrolling"),
         info_students = paste(n_students_fmt, "Kindergarteners")) →
  aux_info
```

A little kludge

```
1  ## This is not an efficient way to do this
2  aux_info
```

```
# A tibble: 11 × 7
   mwc            n_schools
n_students n_schools_fmt
n_students_fmt info_schools
   <fct>              <int>
<dbl> <chr>          <chr>
<chr>
 1 Public              5314
472802 5,314          472,802
5,314 Schoo…
 2 Charter              314
19863 314            19,863
314 Schools…
 3 Private Non-S…       591
16697 591            16,697
591 Schools…
 4 Private Chris…       336
```

A little kludge

```
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students)
```

```
# A tibble: 11 × 3
   mwc
info_schools            info_students
   <fct>                        <chr>
<chr>
 1 Public                       5,314
Schools Enrolling 472,802
Kindergarteners
 2 Charter                      314
Schools Enrolling   19,863
Kindergarteners
 3 Private Non-Specific         591
Schools Enrolling   16,697
Kindergarteners
 4 Private Christian            336
Schools Enrolling   8,836
Kindergarteners
```

A little kludge

```
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character))
```

```
# A tibble: 11 × 3
   mwc
info_schools            info_students
   <chr>                        <chr>
<chr>
 1 Public                       5,314
Schools Enrolling 472,802
Kindergarteners
 2 Charter                      314
Schools Enrolling   19,863
Kindergarteners
 3 Private Non-Specific         591
Schools Enrolling   16,697
Kindergarteners
 4 Private Christian            336
Schools Enrolling   8,836
Kindergarteners
```

A little kludge

```
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_by(mwc)
```

```
# A tibble: 11 × 3
# Groups:   mwc [11]
   mwc
info_schools            info_students
   <chr>                        <chr>
<chr>
 1 Public                       5,314
Schools Enrolling 472,802
Kindergarteners
 2 Charter                      314
Schools Enrolling   19,863
Kindergarteners
 3 Private Non-Specific         591
Schools Enrolling   16,697
Kindergarteners
 4 Private Christian            336
Schools Enrolling   8,836
```

A little kludge

```
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_by(mwc) ▷
  group_keys()
```

```
# A tibble: 11 × 1
   mwc
   <chr>
 1 Charter
 2 Charter Montessori
 3 Private Catholic
 4 Private Christian
 5 Private Christian Montessori
 6 Private Jewish/Islamic
 7 Private Montessori
 8 Private Non-Specific
 9 Private Waldorf
10 Public
11 Public Montessori
```

A little kludge

```
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_by(mwc) ▷
  group_keys() ▷
  pull()
```

```
 [1] "Charter"
"Charter Montessori"
 [3] "Private Catholic"
"Private Christian"
 [5] "Private Christian Montessori"
"Private Jewish/Islamic"
 [7] "Private Montessori"
"Private Non-Specific"
 [9] "Private Waldorf"
"Public"
[11] "Public Montessori"
```

A little kludge

```
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_by(mwc) ▷
  group_keys() ▷
  pull() ▷
  as.character()
```

```
 [1] "Charter"
"Charter Montessori"
 [3] "Private Catholic"
"Private Christian"
 [5] "Private Christian Montessori"
"Private Jewish/Islamic"
 [7] "Private Montessori"
"Private Non-Specific"
 [9] "Private Waldorf"
"Public"
[11] "Public Montessori"
```

A little kludge

```
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_by(mwc) ▷
  group_keys() ▷
  pull() ▷
  as.character() →
  keys
```

A little kludge

```
## This is not an efficient way to do this
aux_info |>
  select(mwc, info_schools, info_students) |>
  mutate(across(everything(), as.character)) |>
  group_by(mwc) |>
  group_keys() |>
  pull() |>
  as.character() ->
  keys

aux_info
```

```
# A tibble: 11 × 7
   mwc            n_schools
n_students n_schools_fmt
n_students_fmt info_schools
   <fct>              <int>
<dbl> <chr>          <chr>
<chr>
 1 Public               5314
472802 5,314          472,802
5,314 Schoo…
 2 Charter               314
19863 314             19,863
314 Schools…
 3 Private Non-S…       591
16697 591             16,697
591 Schools…
 4 Private Chris…       336
```

A little kludge

```
## This is not an efficient way to do this
aux_info |>
  select(mwc, info_schools, info_students) |>
  mutate(across(everything(), as.character)) |>
  group_by(mwc) |>
  group_keys() |>
  pull() |>
  as.character() ->
  keys

aux_info |>
  select(mwc, info_schools, info_students)
```

```
# A tibble: 11 × 3
   mwc
info_schools            info_students
   <fct>                        <chr>
<chr>
 1 Public                       5,314
Schools Enrolling 472,802
Kindergarteners
 2 Charter                      314
Schools Enrolling   19,863
Kindergarteners
 3 Private Non-Specific         591
Schools Enrolling   16,697
Kindergarteners
 4 Private Christian            336
Schools Enrolling   8,836
Kindergarteners
```

A little kludge

```
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_by(mwc) ▷
  group_keys() ▷
  pull() ▷
  as.character() →
  keys

aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character))
```

```
# A tibble: 11 × 3
   mwc
info_schools            info_students
   <chr>                        <chr>
<chr>
 1 Public                       5,314
Schools Enrolling 472,802
Kindergarteners
 2 Charter                      314
Schools Enrolling   19,863
Kindergarteners
 3 Private Non-Specific         591
Schools Enrolling   16,697
Kindergarteners
 4 Private Christian            336
Schools Enrolling   8,836
Kindergarteners
```

A little kludge

```r
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_by(mwc) ▷
  group_keys() ▷
  pull() ▷
  as.character() →
  keys

aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_split(mwc)
```

```
<list_of<
  tbl_df<
    mwc          : character
    info_schools : character
    info_students: character
  >
>[11]>
[[1]]
# A tibble: 1 × 3
  mwc     info_schools
info_students
  <chr>   <chr>                  <chr>
1 Charter 314 Schools Enrolling
19,863 Kindergarteners

[[2]]
# A tibble: 1 × 3
```

A little kludge

```
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_by(mwc) ▷
  group_keys() ▷
  pull() ▷
  as.character() →
  keys

aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_split(mwc) ▷
  set_names(keys)   # There's a better way ...
```

```
<list_of<
  tbl_df<
    mwc          : character
    info_schools : character
    info_students: character
  >
>[11]>
$Charter
# A tibble: 1 × 3
  mwc     info_schools
info_students
  <chr>   <chr>                  <chr>
1 Charter 314 Schools Enrolling
19,863 Kindergarteners

$`Charter Montessori`
# A tibble: 1 × 3
```

A little kludge

```r
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_by(mwc) ▷
  group_keys() ▷
  pull() ▷
  as.character() →
  keys

aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_split(mwc) ▷
  set_names(keys) ▷  # There's a better way ...
  map_chr(.f = paste, sep = "", collapse = "\n")
```

```
Charter
                "Charter\n314 Schools
Enrolling\n19,863 Kindergarteners"

Charter Montessori
          "Charter Montessori\n5
Schools Enrolling\n227
Kindergarteners"

Private Catholic
        "Private Catholic\n334 Schools
Enrolling\n9,869 Kindergarteners"

Private Christian
       "Private Christian\n336 Schools
Enrolling\n8,836 Kindergarteners"
```

A little kludge

```
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_by(mwc) ▷
  group_keys() ▷
  pull() ▷
  as.character() →
  keys

aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_split(mwc) ▷
  set_names(keys) ▷  # There's a better way ...
  map_chr(.f = paste, sep = "", collapse = "\n") →
```

A little kludge

```
## This is not an efficient way to do this
aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_by(mwc) ▷
  group_keys() ▷
  pull() ▷
  as.character() →
  keys

aux_info ▷
  select(mwc, info_schools, info_students) ▷
  mutate(across(everything(), as.character)) ▷
  group_split(mwc) ▷
  set_names(keys) ▷  # There's a better way ...
  map_chr(.f = paste, sep = "", collapse = "\n") →
```

At last, the Beeplot

```
1 cavax
```

```
# A tibble: 7,032 × 13
     code county name  type  district
city  enrollment pbe_pct exempt
med_exempt
    <dbl> <chr>  <chr> <chr> <chr>
<chr>       <dbl>   <dbl>  <dbl>
<dbl>
 1 1.10e5 ALAME… FAME… PUBL… ALAMEDA…
NEWA…         109       13  12.8
0
 2 6.00e6 ALAME… COX … PUBL… ALAMEDA…
OAKL…         115        1   0.87
0.87
 3 6.00e6 ALAME… LAZE… PUBL… ALAMEDA…
OAKL…          40        0   0
0
 4 1.24e5 ALAME… YU M… PUBL… ALAMEDA…
```

At last, the Beeplot

```
cavax ▷
  filter(mwc %nin% c("Private Christian Montessori",
                     "Charter Montessori",
                     "Private Jewish/Islamic"))
```

```
# A tibble: 7,015 × 13
     code county name  type  district
city  enrollment pbe_pct exempt
med_exempt
    <dbl> <chr>  <chr> <chr> <chr>
<chr>       <dbl>   <dbl>  <dbl>
<dbl>
 1 1.10e5 ALAME… FAME… PUBL… ALAMEDA…
NEWA…         109      13  12.8
0
 2 6.00e6 ALAME… COX … PUBL… ALAMEDA…
OAKL…         115       1   0.87
0.87
 3 6.00e6 ALAME… LAZE… PUBL… ALAMEDA…
OAKL…          40       0   0
0
 4 1.24e5 ALAME… YU M… PUBL… ALAMEDA…
```

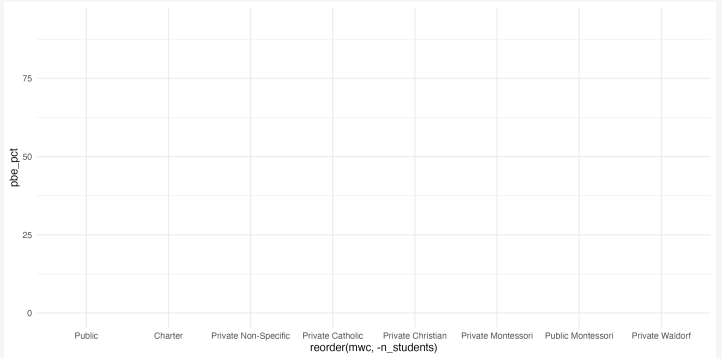
At last, the Beeplot

```
cavax ▷
  filter(mwc %nin% c("Private Christian Montessori",
                     "Charter Montessori",
                     "Private Jewish/Islamic")) ▷
  left_join(aux_info, by = "mwc")
```

```
# A tibble: 7,015 × 19
     code county name  type  district
city  enrollment pbe_pct exempt
med_exempt
    <dbl> <chr>  <chr> <chr> <chr>
<chr>       <dbl>   <dbl>  <dbl>
<dbl>
 1 1.10e5 ALAME… FAME… PUBL… ALAMEDA…
NEWA…         109      13  12.8
0
 2 6.00e6 ALAME… COX … PUBL… ALAMEDA…
OAKL…         115       1   0.87
0.87
 3 6.00e6 ALAME… LAZE… PUBL… ALAMEDA…
OAKL…          40       0   0
0
 4 1.24e5 ALAME… YU M… PUBL… ALAMEDA…
```

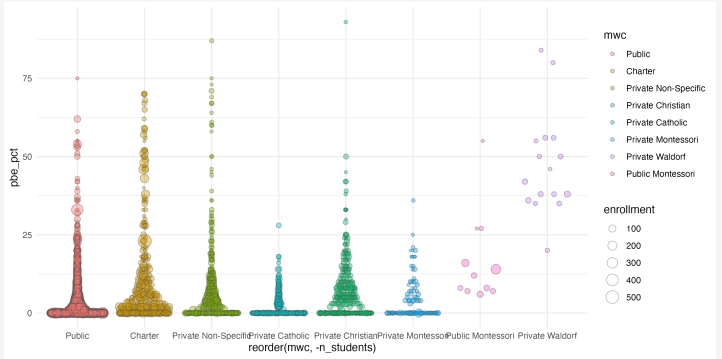
At last, the Beeplot

```
cavax ▷
  filter(mwc %nin% c("Private Christian Montessori",
                     "Charter Montessori",
                     "Private Jewish/Islamic")) ▷
  left_join(aux_info, by = "mwc") ▷
  ggplot(mapping =
           aes(y = pbe_pct,
               x = reorder(mwc, -n_students),
               size = enrollment,
               fill = mwc))
```



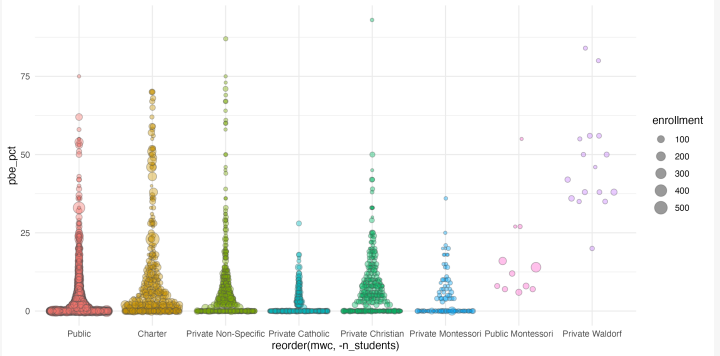
At last, the Beeplot

```
cavax ▷
  filter(mwc %nin% c("Private Christian Montessori",
                     "Charter Montessori",
                     "Private Jewish/Islamic")) ▷
  left_join(aux_info, by = "mwc") ▷
  ggplot(mapping =
           aes(y = pbe_pct,
               x = reorder(mwc, -n_students),
               size = enrollment,
               fill = mwc)) +
  geom_quasirandom(shape=21,
        alpha = 0.4,color="gray30",
        method = "quasirandom",
        varwidth = FALSE,
        bandwidth = 0.9)
```



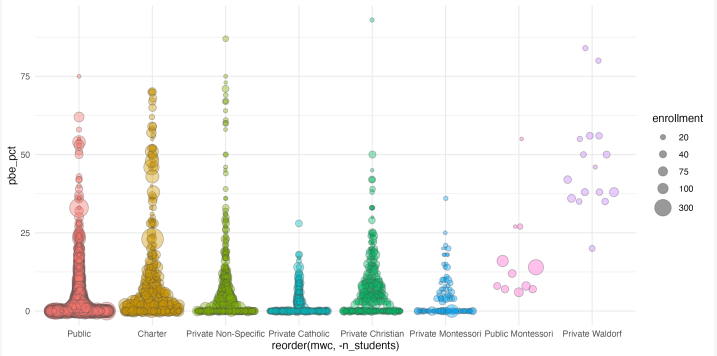
At last, the Beeplot

```
cavax ▷
  filter(mwc %nin% c("Private Christian Montessori",
                     "Charter Montessori",
                     "Private Jewish/Islamic")) ▷
  left_join(aux_info, by = "mwc") ▷
  ggplot(mapping =
           aes(y = pbe_pct,
               x = reorder(mwc, -n_students),
               size = enrollment,
               fill = mwc)) +
  geom_quasirandom(shape=21,
        alpha = 0.4,color="gray30",
        method = "quasirandom",
        varwidth = FALSE,
        bandwidth = 0.9) +
      guides(color = "none",
```



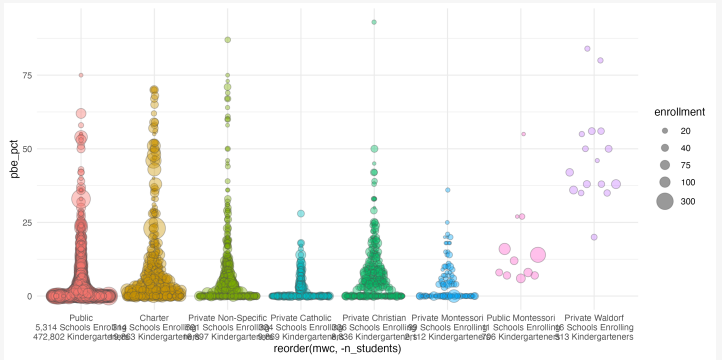
At last, the Beeplot

```
cavax ▷
  filter(mwc %nin% c("Private Christian Montessori",
                     "Charter Montessori",
                     "Private Jewish/Islamic")) ▷
  left_join(aux_info, by = "mwc") ▷
  ggplot(mapping =
           aes(y = pbe_pct,
               x = reorder(mwc, -n_students),
               size = enrollment,
               fill = mwc)) +
  geom_quasirandom(shape=21,
        alpha = 0.4,color="gray30",
        method = "quasirandom",
        varwidth = FALSE,
        bandwidth = 0.9) +
      guides(color = "none",
```



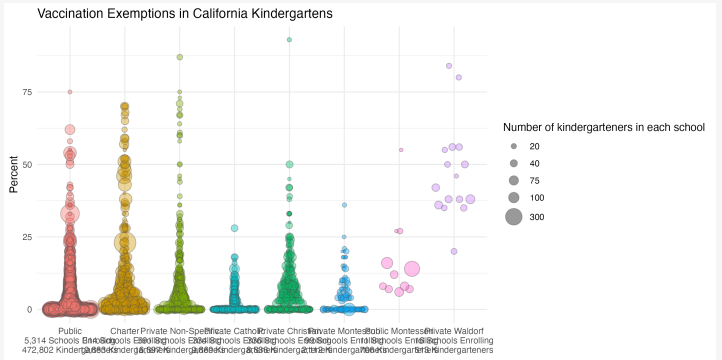
At last, the Beeplot

```
  filter(mwc %nin% c("Private Christian Montessori",
                     "Charter Montessori",
                     "Private Jewish/Islamic")) ▷
  left_join(aux_info, by = "mwc") ▷
  ggplot(mapping =
           aes(y = pbe_pct,
               x = reorder(mwc, -n_students),
               size = enrollment,
               fill = mwc)) +
  geom_quasirandom(shape=21,
        alpha = 0.4,color="gray30",
        method = "quasirandom",
        varwidth = FALSE,
        bandwidth = 0.9) +
      guides(color = "none",
          shape= "none",
```



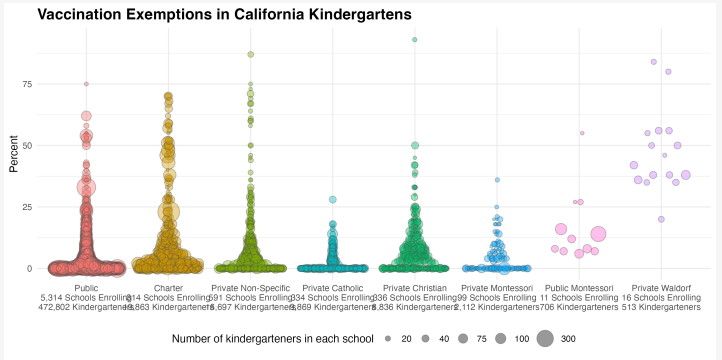
At last, the Beeplot

```
 5   left_join(aux_info, by = "mwc") ▷
 6   ggplot(mapping =
 7            aes(y = pbe_pct,
 8                x = reorder(mwc, -n_students),
 9                size = enrollment,
10                fill = mwc)) +
11   geom_quasirandom(shape=21,
12         alpha = 0.4,color="gray30",
13         method = "quasirandom",
14         varwidth = FALSE,
15         bandwidth = 0.9) +
16       guides(color = "none",
17           shape= "none",
18           fill= "none",
19           size = guide_legend(override.aes =
20                      list(fill = "black"))) +
```



At last, the Beeplot

```
8                x = reorder(mwc, -n_students),
9                size = enrollment,
10               fill = mwc)) +
11   geom_quasirandom(shape=21,
12         alpha = 0.4,color="gray30",
13         method = "quasirandom",
14         varwidth = FALSE,
15         bandwidth = 0.9) +
16       guides(color = "none",
17           shape= "none",
18           fill= "none",
19           size = guide_legend(override.aes =
20                      list(fill = "black"))) +
21       scale_size(breaks=c(20, 40, 75, 100, 300),
22                  range=c(1,10)) +
23       scale_x_discrete(labels = special_x_labs) +
```

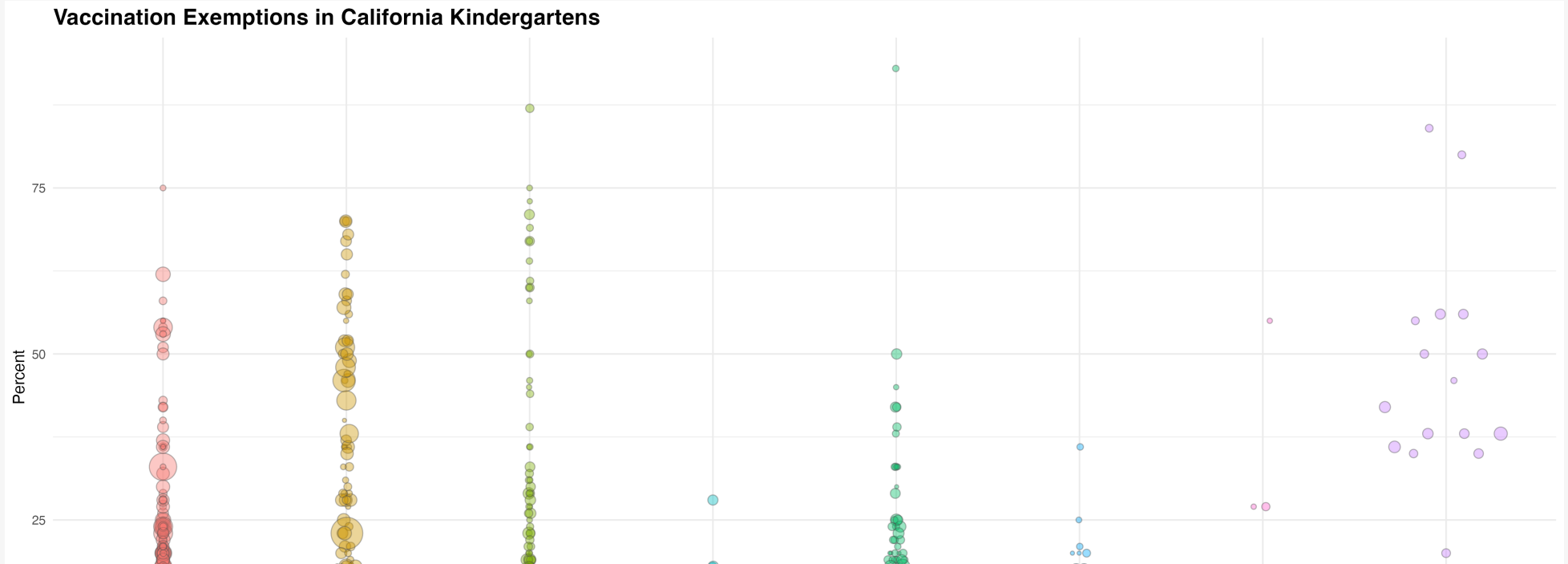


At last, the Beeplot

```
               size = enrollment,
               fill = mwc)) +
  geom_quasirandom(shape=21,
        alpha = 0.4,color="gray30",
        method = "quasirandom",
        varwidth = FALSE,
        bandwidth = 0.9) +
      guides(color = "none",
          shape= "none",
          fill= "none",
          size = guide_legend(override.aes =
                    list(fill = "black"))) +
      scale_size(breaks=c(20, 40, 75, 100, 300),
                 range=c(1,10)) +
      scale_x_discrete(labels = special_x_labs) +
      labs(size = "Number of kindergarteners in each
```

At last, the Beeplot

```
              size = enrollment,
              fill = mwc)) +
  geom_quasirandom(shape=21,
        alpha = 0.4,color="gray30",
        method = "quasirandom",
        varwidth = FALSE,
        bandwidth = 0.9) +
      guides(color = "none",
          shape= "none",
          fill= "none",
          size = guide_legend(override.aes =
                    list(fill = "black"))) +
      scale_size(breaks=c(20, 40, 75, 100, 300),
                 range=c(1,10)) +
      scale_x_discrete(labels = special_x_labs) +
      labs(size = "Number of kindergarteners in each
```


Vaccination Exemptions in California Kindergartens
Percent
75
50
25