Plot My-Sample

Contents


```
metaviz_long <- rio::import(here::here("data", "metaviz_long.rds"))
```

Explore Data for Gene Analysis

All variables

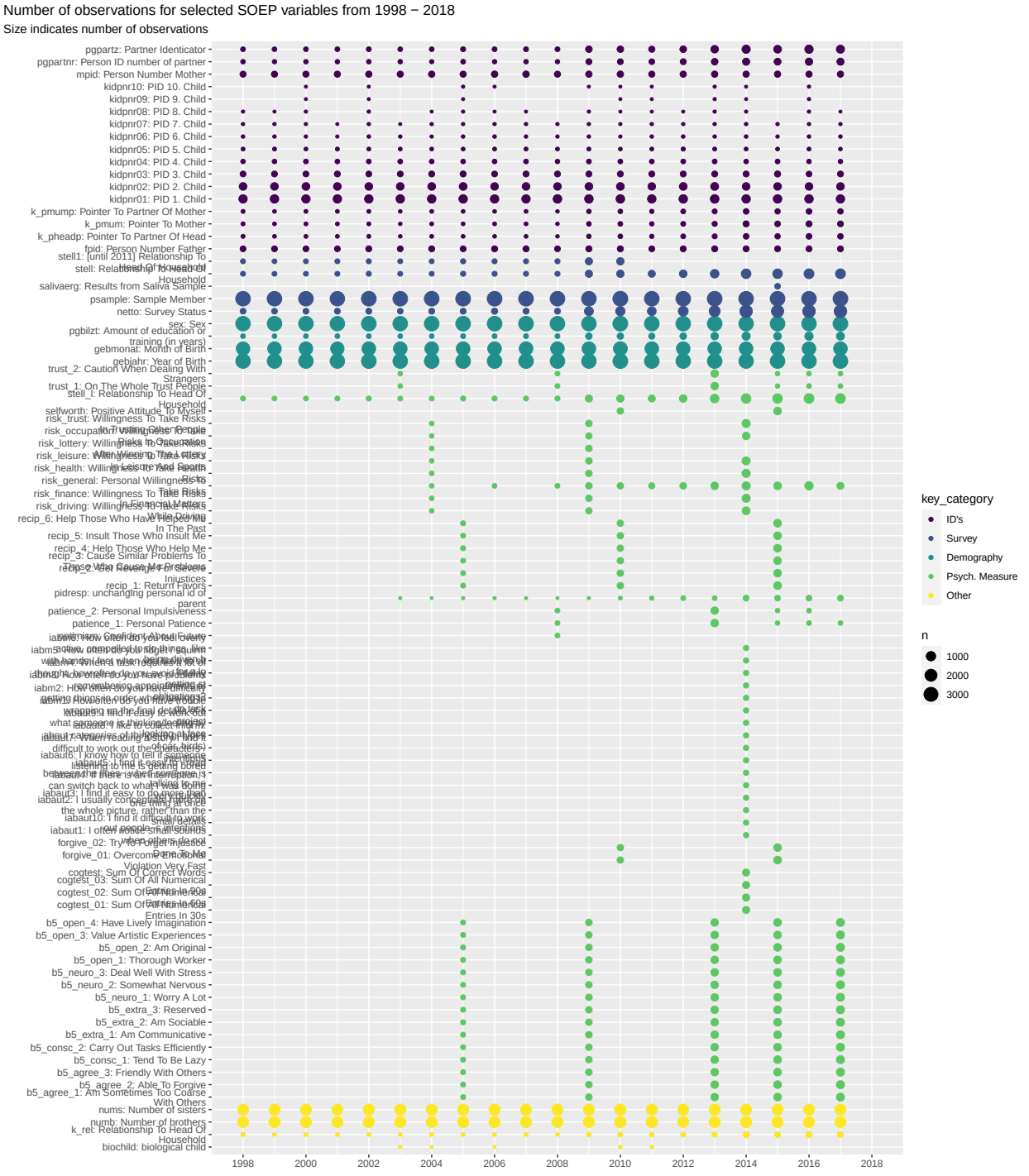
By Year

First you can see the number of available observation for each variable in each year

- x-axis = survey year
- y-axis = variables
- size = number of observations
- colour = variable group

```
metaviz_long %>%
      drop_na(value) %>%
      mutate(key_name_label = factor(key_name_label),
             order = as.numeric(key_category),
             key_name_label = fct_reorder(key_name_label, desc(order))) %>%
      ggplot(aes(key_name_label, syear, col = key_category)) +
      geom_count() +
      coord_flip() +
      theme(legend.position="right",
           plot.title.position = "plot") + #so cool <3)
      guides(col = guide_legend(ncol = 1)) +
      scale_x_discrete(labels = wrap_format(40))+
      scale_y_continuous(limits= c(1998, 2018), breaks = seq(1998,2018,2))+
```

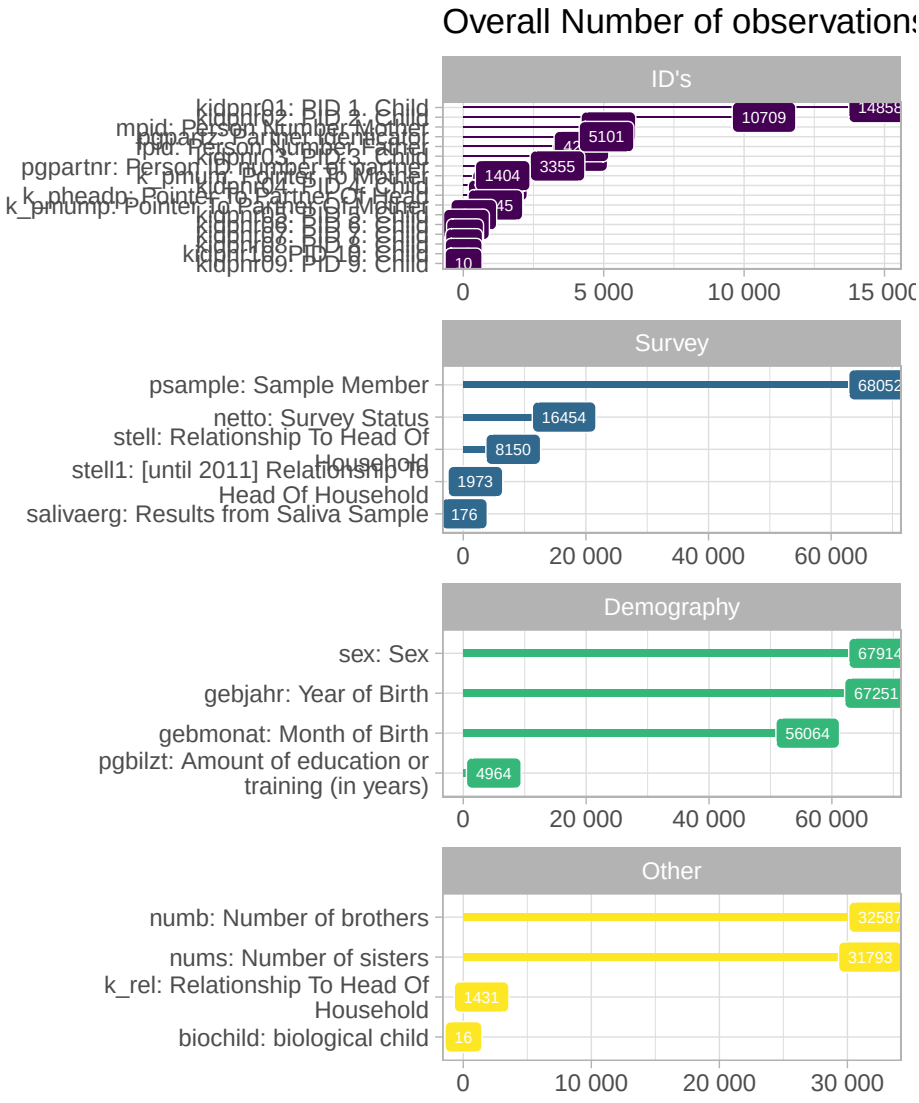
```
    labs(title = "Number of observations for selected SOEP variables from 1998 - 2018",
         subtitle = "Size indicates number of observations",
         y = "", x = "")
```



Overall

Here is an overall plot of the number of available observations for each of the variables. It helps to get a general understanding of the proportions of missings for groups of variables

```
metaviz_long %>%
      drop_na(value) %>%
      filter(key_category != "Psych. Measure") %>%
      group_by(key) %>%
      add_count() %>%
      ungroup() %>%
      distinct(key, .keep_all = T) %>%
      group_by(key_category) %>%
      mutate(key_name_label = fct_reorder(factor(key_name_label), n)) %>%
      ggplot(aes(x = key_name_label, y = n, fill = key_category, label = n)) +
      geom_col(width = 0.2) +
      geom_point() +
      geom_label(color = "white", size = 2) +
      coord_flip() +
      scale_y_continuous(labels = scales::label_number_auto()) +
      scale_x_discrete(labels = wrap_format(40))+
      theme_light() +
      theme(legend.position = "none") +
      facet_wrap(~key_category, ncol = 1, scales = "free") +
      labs(title = "Overall Number of observations for selected SOEP variables from 1998 - 2018", y = "
```

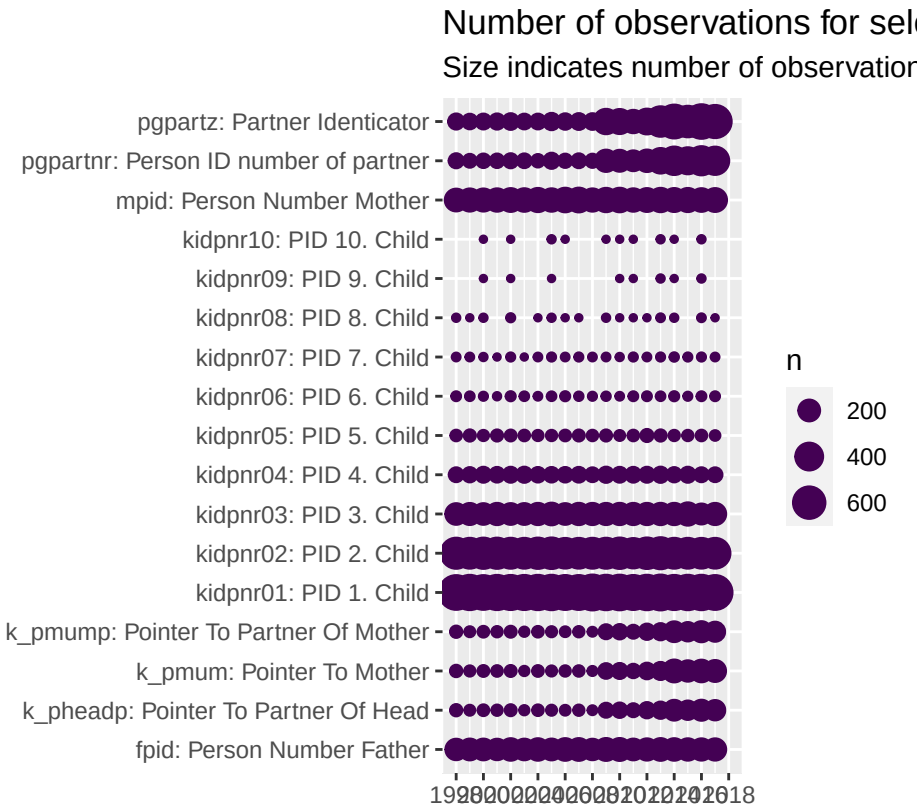


By Variable Category {.tabset}

By Year

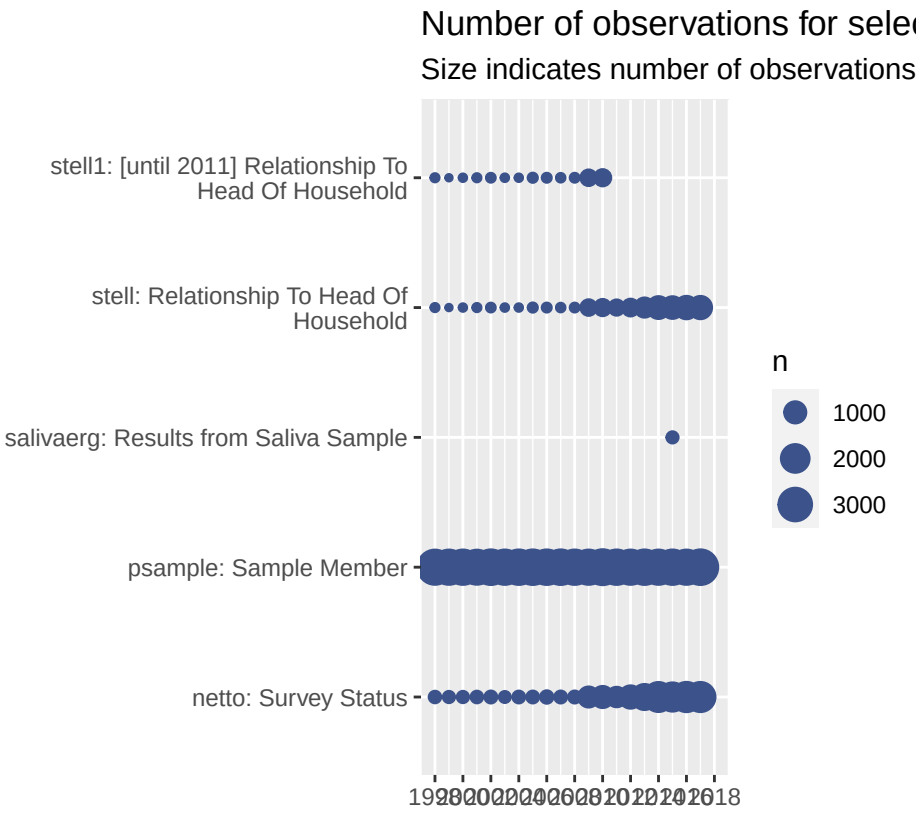
```
metaviz_long %>%
      drop_na(value) %>%
      filter(key_category == "ID's") %>%
      ggplot(aes(key_name_label, syear)) +
      geom_count(col= "#440154FF") +
      coord_flip() +
      theme(legend.position="right") +
      scale_x_discrete(labels = wrap_format(40))+
      scale_y_continuous(limits= c(1998, 2018), breaks = seq(1998,2018,2)) +
      labs(title = "Number of observations for selected SOEP variables from 1998 - 2018",
```

```
          subtitle = "Size indicates number of observations",
          y = "", x = "")
```



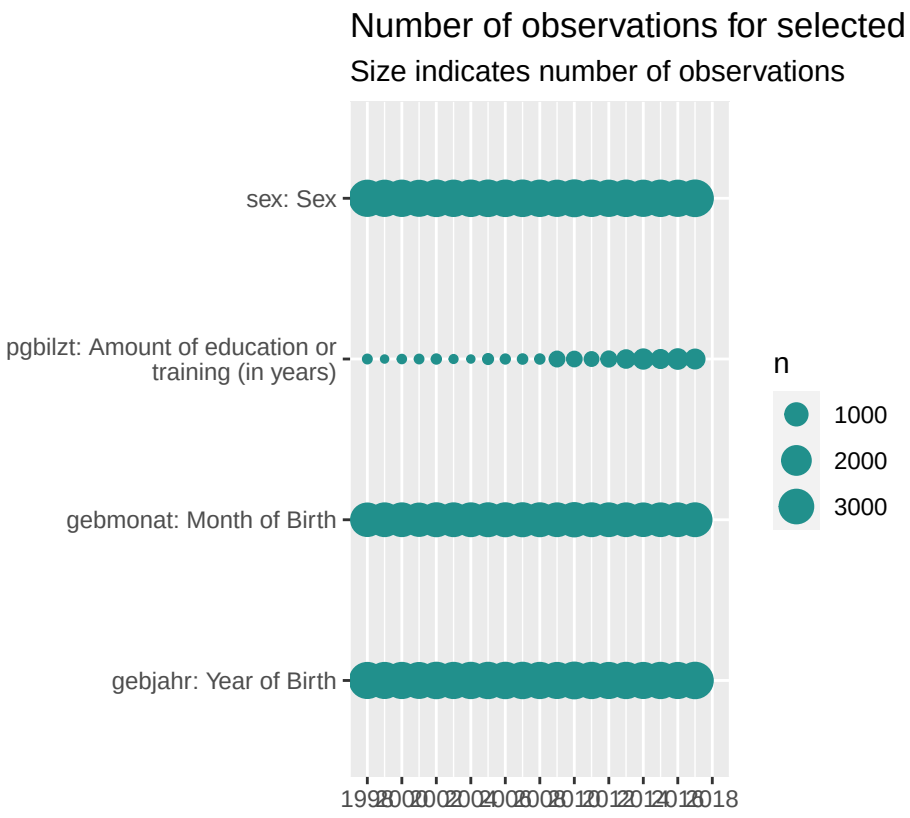
ID's

```
metaviz_long %>%
      drop_na(value) %>%
      filter(key_category == "Survey") %>%
      ggplot(aes(key_name_label, syear)) +
      geom_count(col= "#3B528BFF") +
      coord_flip() +
      theme(legend.position="right") +
      scale_x_discrete(labels = wrap_format(40))+
      scale_y_continuous(limits= c(1998, 2018), breaks = seq(1998,2018,2)) +
      labs(title = "Number of observations for selected SOEP variables from 1998 - 2018",
           subtitle = "Size indicates number of observations",
           y = "", x = "")
```



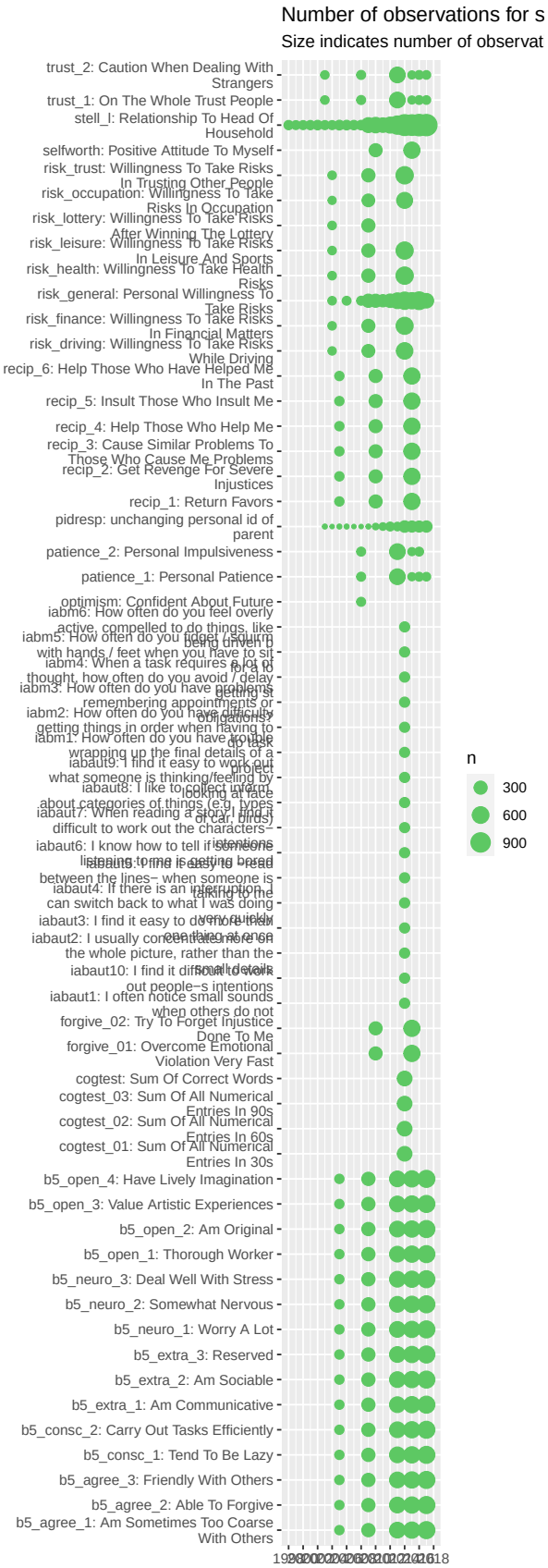
Survey

```
metaviz_long %>%
      drop_na(value) %>%
      filter(key_category == "Demography") %>%
      ggplot(aes(key_name_label, syear)) +
      geom_count(col= "#21908CFF") +
      coord_flip() +
      theme(legend.position="right") +
      scale_x_discrete(labels = wrap_format(40))+
      scale_y_continuous(limits= c(1998, 2018), breaks = seq(1998,2018,2)) +
      labs(title = "Number of observations for selected SOEP variables from 1998 - 2018",
           subtitle = "Size indicates number of observations",
           y = "", x = "")
```



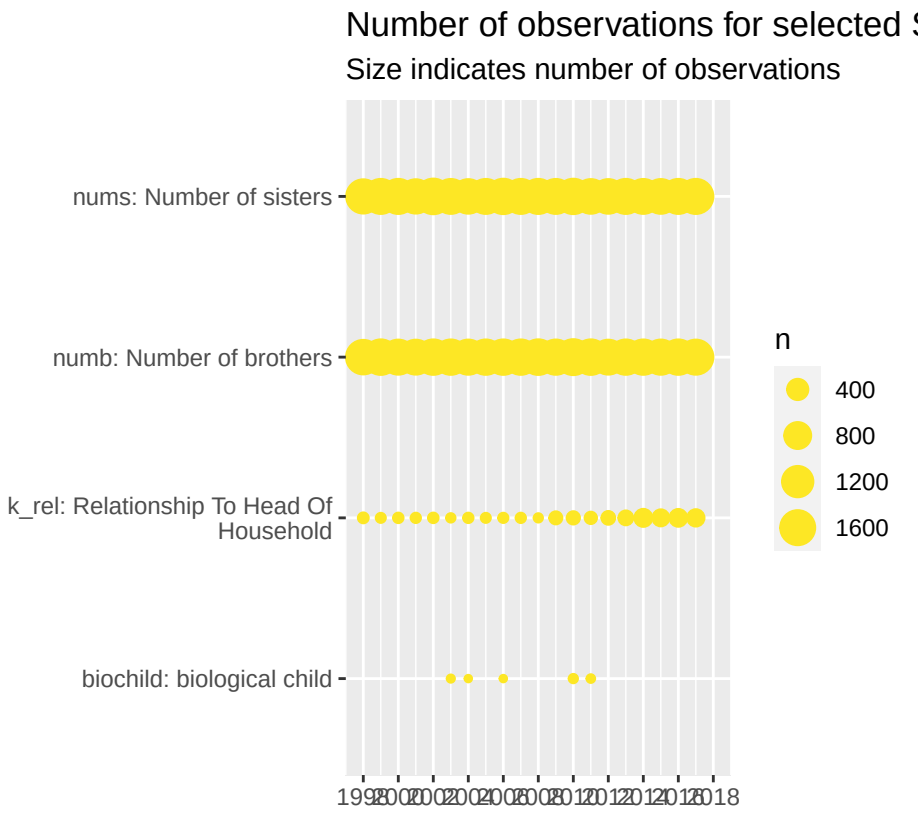
Demography

```
metaviz_long %>%
      drop_na(value) %>%
      filter(key_category == "Psych. Measure") %>%
      ggplot(aes(key_name_label, syear)) +
      geom_count(col = "#5DC863FF") +
      coord_flip() +
      theme(legend.position="right") +
      scale_x_discrete(labels = wrap_format(40))+
      scale_y_continuous(limits= c(1998, 2018), breaks = seq(1998,2018,2)) +
      labs(title = "Number of observations for selected SOEP variables from 1998 - 2018",
           subtitle = "Size indicates number of observations",
           y = "", x = "")
```



Psychol. Measures

```
metaviz_long %>%
      drop_na(value) %>%
      filter(key_category == "Other") %>%
      ggplot(aes(key_name_label, syear)) +
      geom_count(col = "#FDE725FF") +
      coord_flip() +
      theme(legend.position="right") +
      scale_x_discrete(labels = wrap_format(40))+
      scale_y_continuous(limits= c(1998, 2018), breaks = seq(1998,2018,2)) +
      labs(title = "Number of observations for selected SOEP variables from 1998 - 2018",
           subtitle = "Size indicates number of observations",
           y = "", x = "")
```



Other

Supplement

resources

- row names to column: https://stackoverflow.com/questions/29511215/convert-row-names-into-first-column
- age categories: https://ggplot2.tidyverse.org/reference/cut_interval.html
- wrap label names: https://stackoverflow.com/questions/21878974/auto-wrapping-of-labels-via-labeller-label-wrap-in-ggplot2