MATH 118: Notes C

Making plots with ggplot2: barplots

```
library(tidyverse)
```

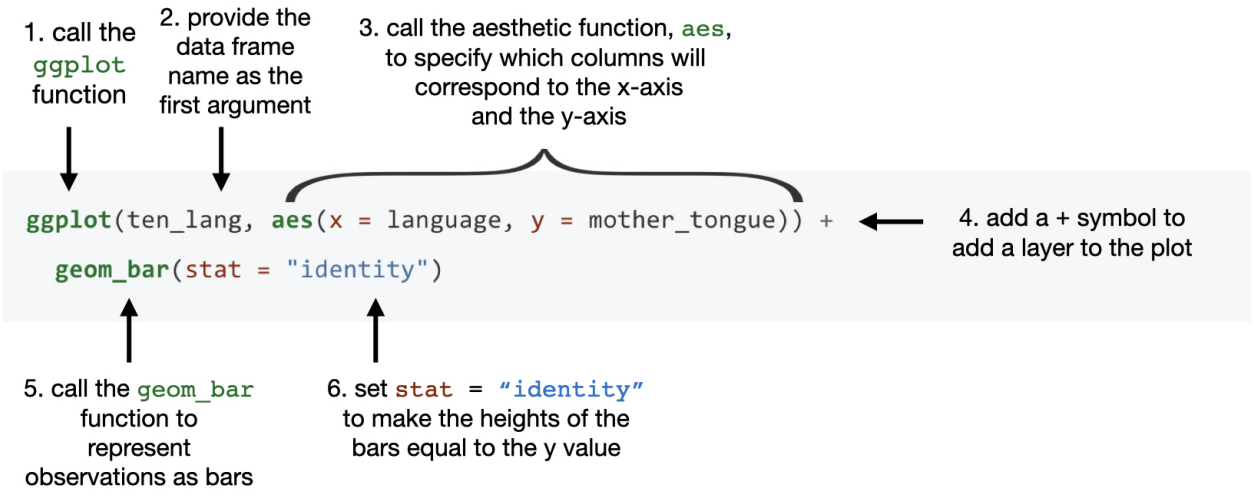
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.7      v dplyr   1.0.10
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#Import the can_lang dataset from last class

#can_lang <- read.csv("can_lang.csv")

#OR
can_lang <- read.csv("https://raw.githubusercontent.com/ttimbers/canlang/master/inst/extdata/can_lang.cs
```

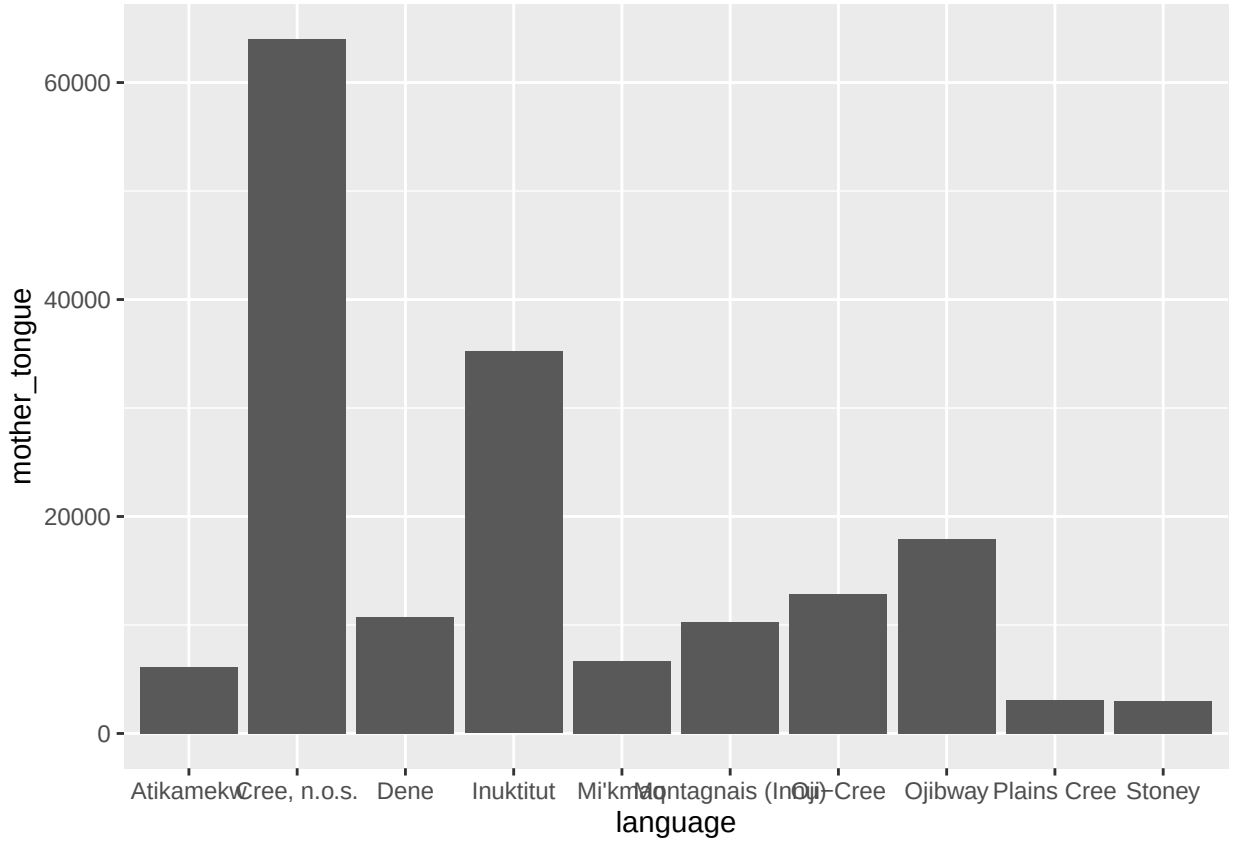
Recall our last example:

```
ten_lang <- can_lang %>%
  filter(category == "Aboriginal languages") %>%
  arrange(by=desc(mother_tongue)) %>%
  select(language, mother_tongue) %>%
  slice(1:10)
```

ggplot: barplots with geom_bar



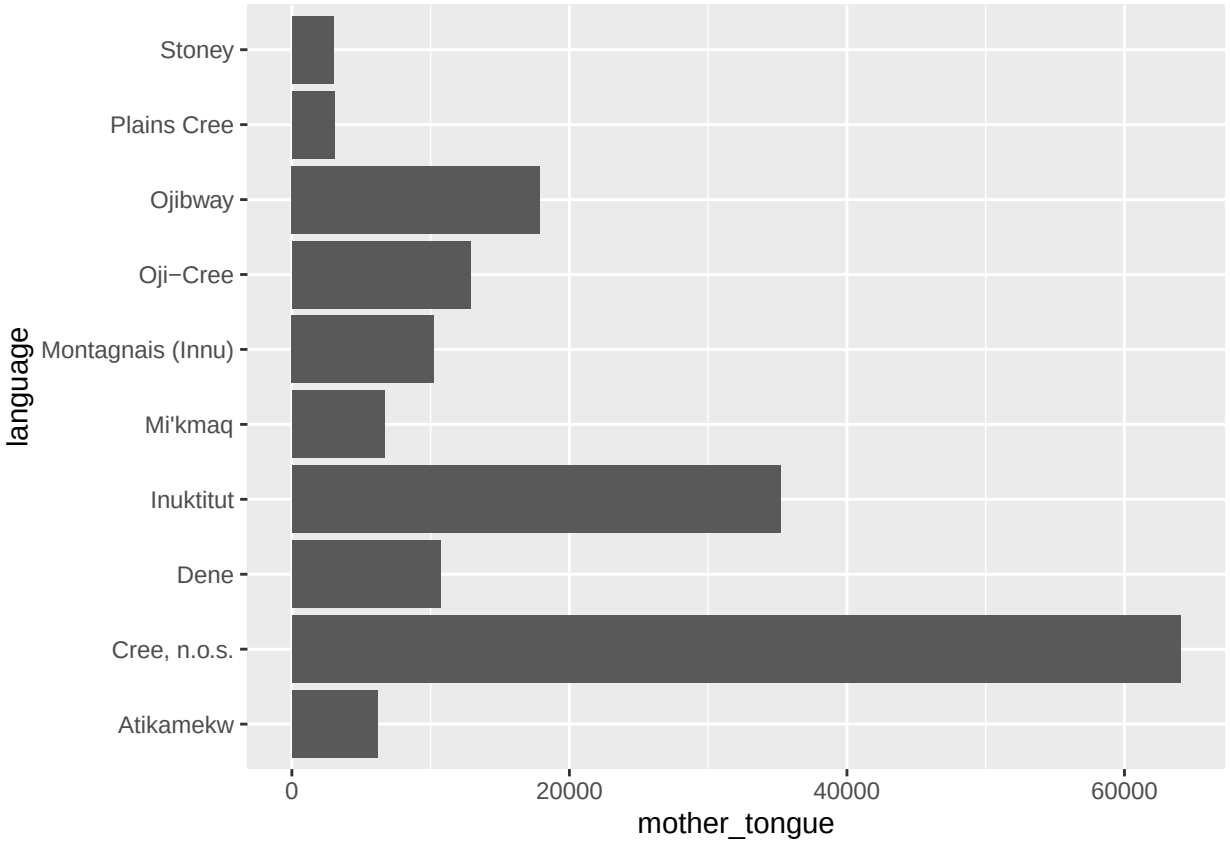

```
ggplot(ten_lang, aes(x = language, y = mother_tongue)) +
  geom_bar(stat = "identity")
```



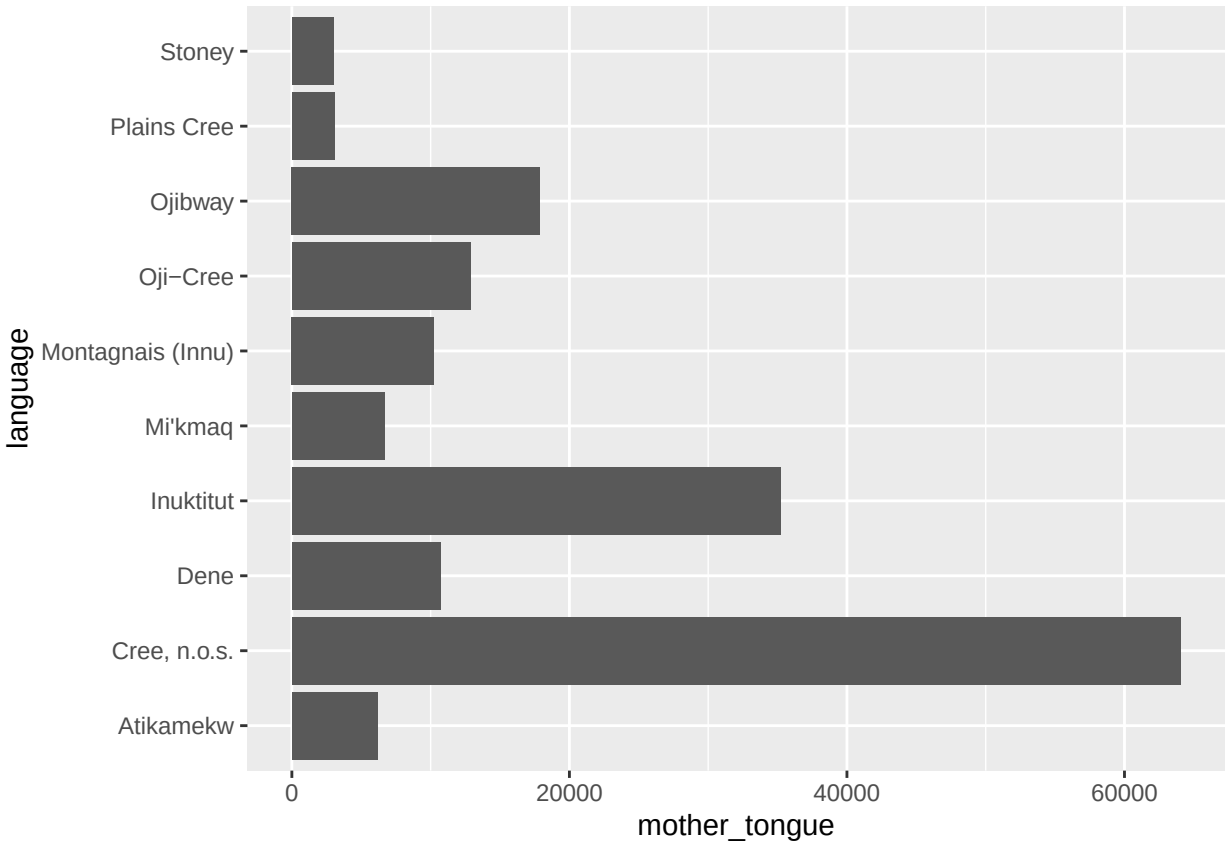
Is there any improvements we could make to this graph?

To better view text

```
ggplot(ten_lang, aes(x = language, y = mother_tongue)) +
  geom_bar(stat = "identity") +
  coord_flip()
```

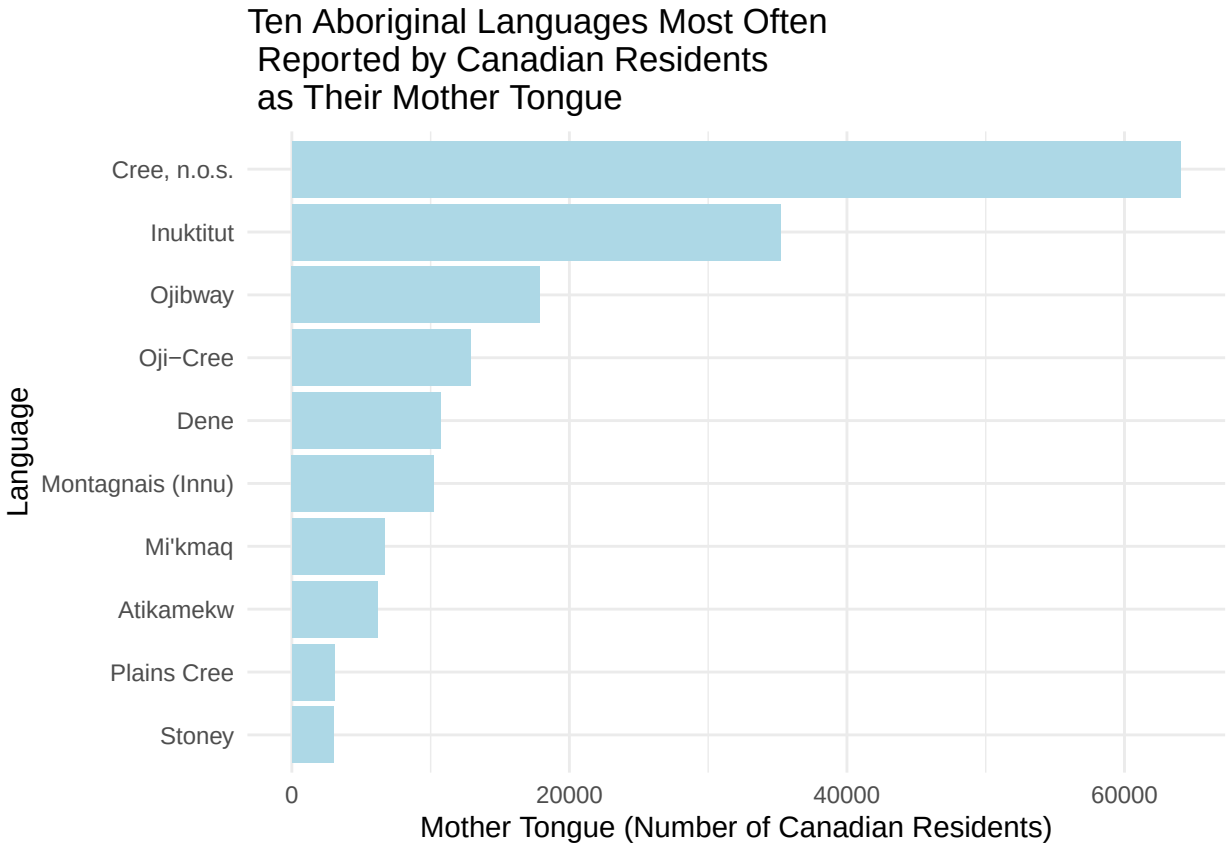



```
#OR

ggplot(ten_lang, aes(x = mother_tongue, y = language)) +
  geom_bar(stat = "identity")
```



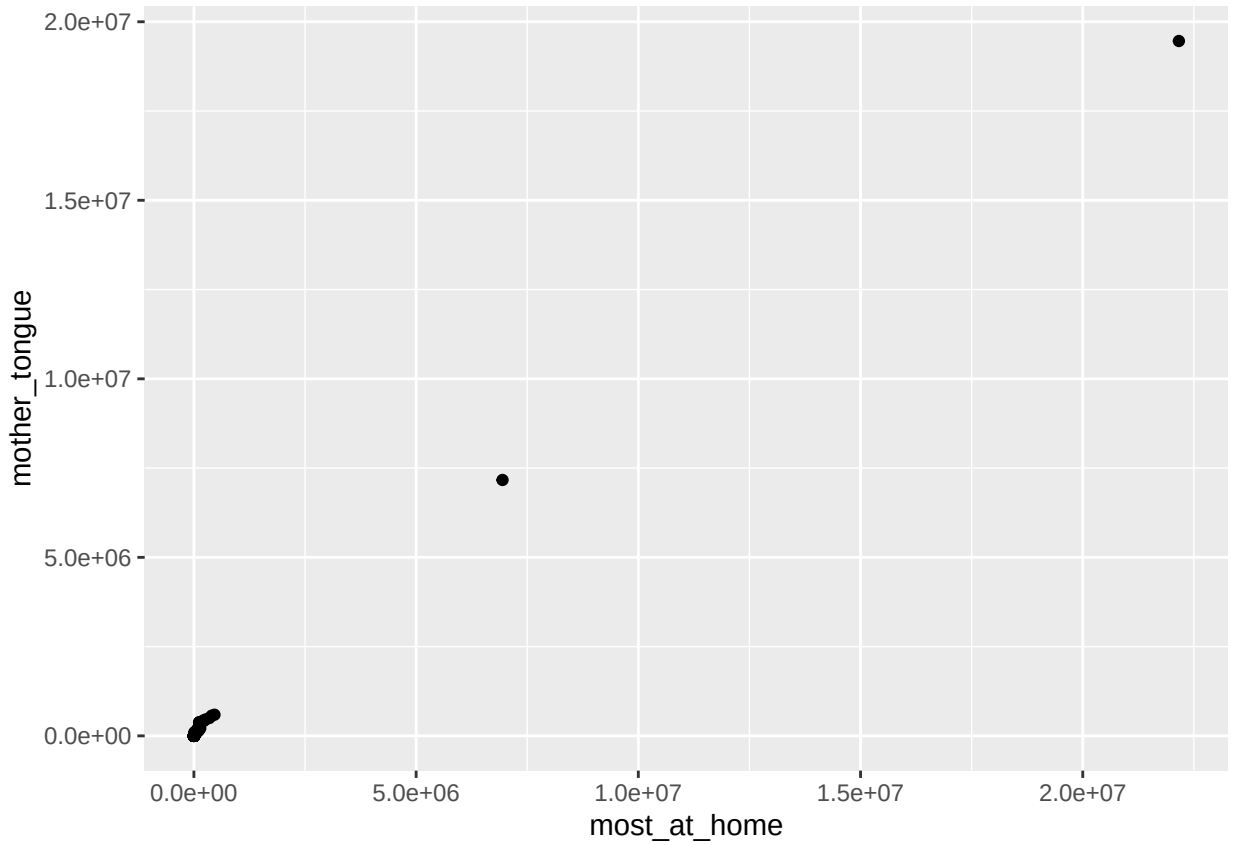
Labels, Colors, and Themes

```
ggplot(ten_lang, aes(x = mother_tongue, y = reorder(language, mother_tongue))) +
  geom_bar(fill="lightblue", stat = "identity") +
  ylab("Language") +
  xlab("Mother Tongue (Number of Canadian Residents)") +
  ggtitle("Ten Aboriginal Languages Most Often \n Reported by Canadian Residents \n as Their Mother Tong
  theme_minimal()
```



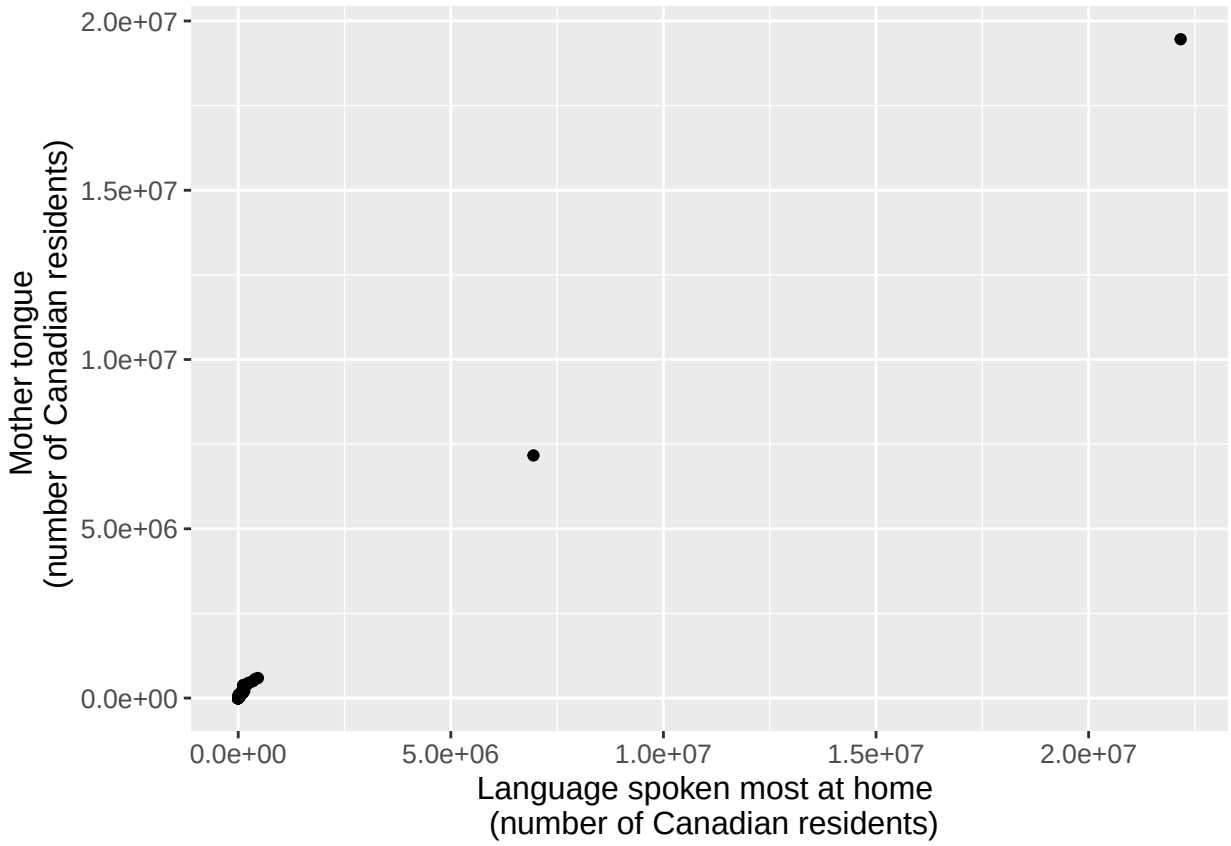
ggplot: scatterplot with `geom_point`

```
ggplot(can_lang, aes(x=most_at_home, y=mother_tongue)) +
  geom_point()
```



With labels

```
ggplot(can_lang, aes(x = most_at_home, y = mother_tongue)) +
  geom_point() +
  xlab("Language spoken most at home \n (number of Canadian residents)") +
  ylab("Mother tongue \n (number of Canadian residents)") +
  theme(text = element_text(size = 12))
```



Axis transformations

```
library(scales)
```

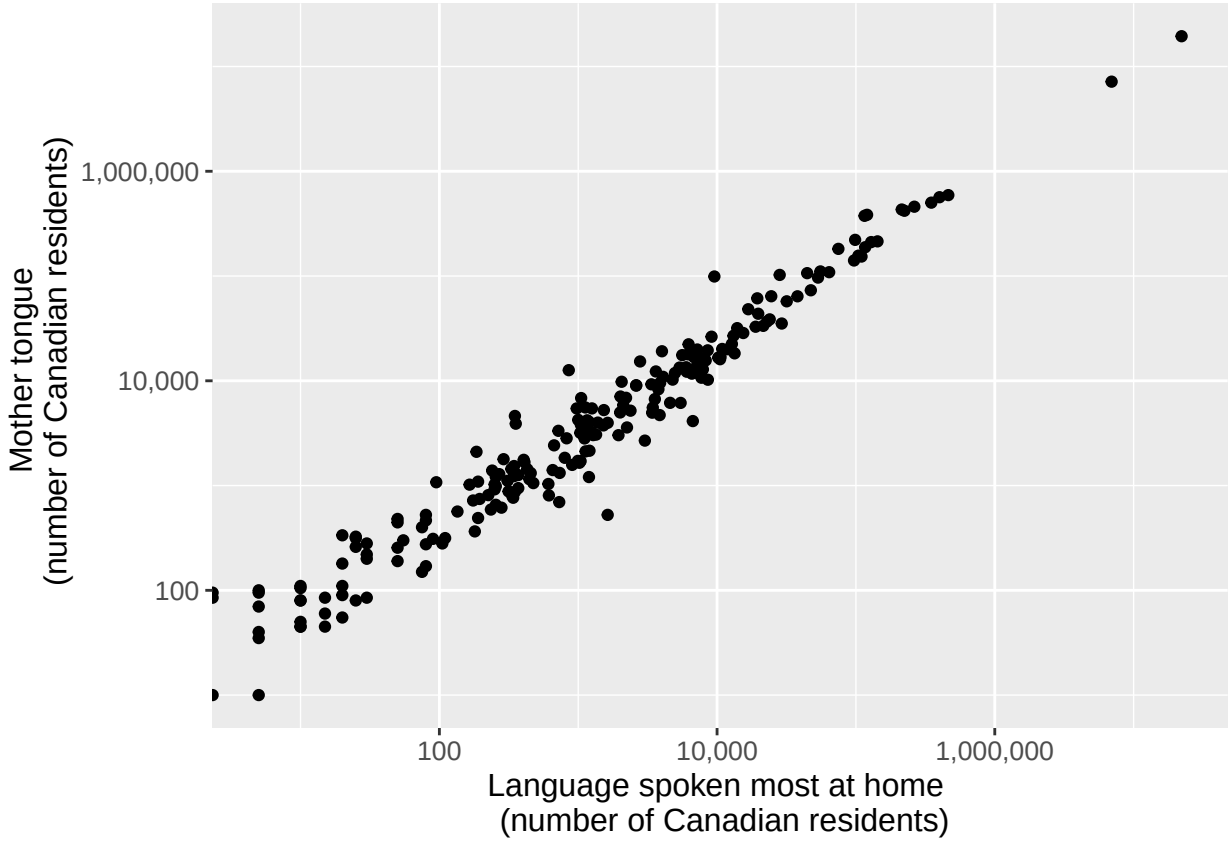
```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
ggplot(can_lang, aes(x = most_at_home, y = mother_tongue)) +
  geom_point() +
  xlab("Language spoken most at home \n (number of Canadian residents)") +
  ylab("Mother tongue \n (number of Canadian residents)") +
  theme(text = element_text(size = 12)) +
  scale_x_log10(labels = label_comma()) +
  scale_y_log10(labels = label_comma())
```

```
## Warning: Transformation introduced infinite values in continuous x-axis
```



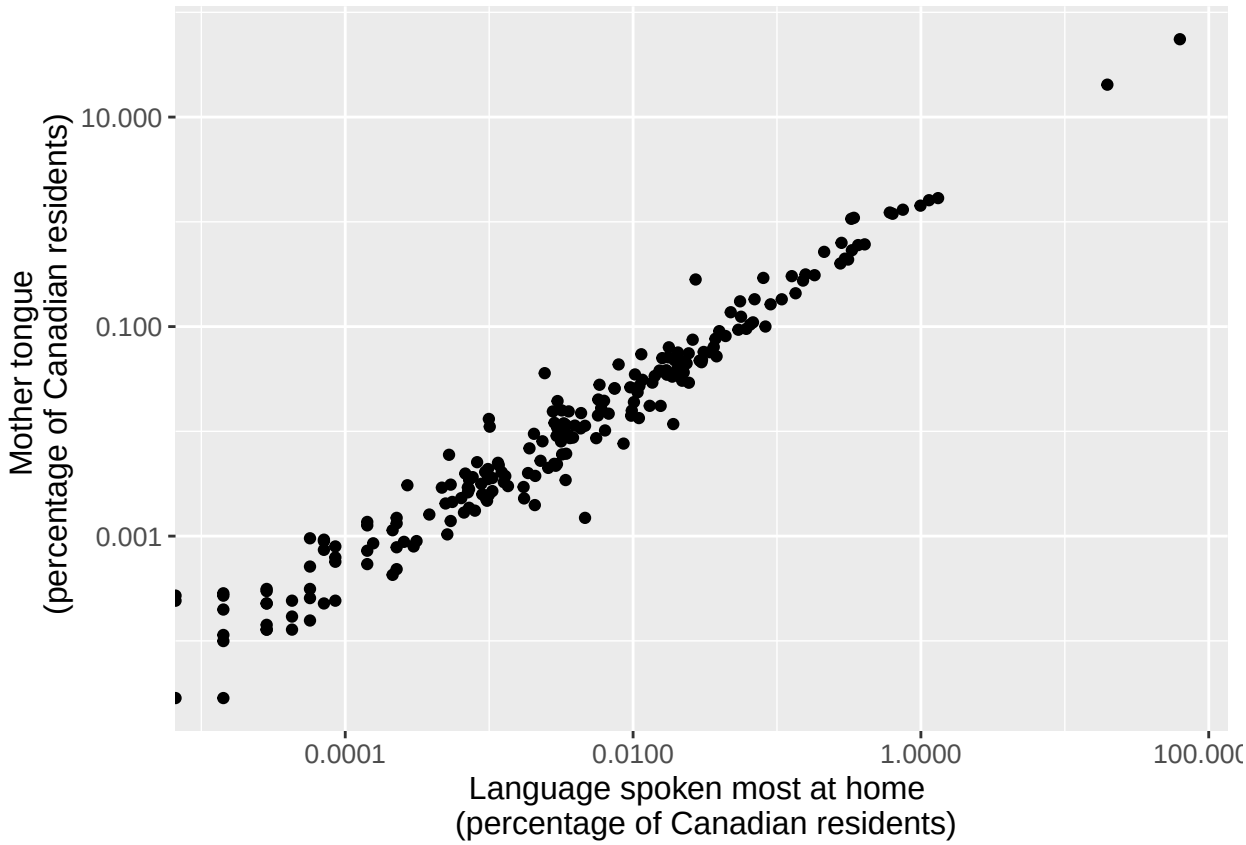
`mutate` to create new columns

```
can_lang <- can_lang %>%
  mutate(
    mother_tongue_percent = (mother_tongue / 35151728) * 100,
    most_at_home_percent = (most_at_home / 35151728) * 100
  )
```

Scatterplot with percents

```
ggplot(can_lang, aes(x = most_at_home_percent, y = mother_tongue_percent)) +
  geom_point() +
  xlab("Language spoken most at home \n (percentage of Canadian residents)") +
  ylab("Mother tongue \n (percentage of Canadian residents)") +
  theme(text = element_text(size = 12)) +
  scale_x_log10(labels = comma) +
  scale_y_log10(labels = comma)
```

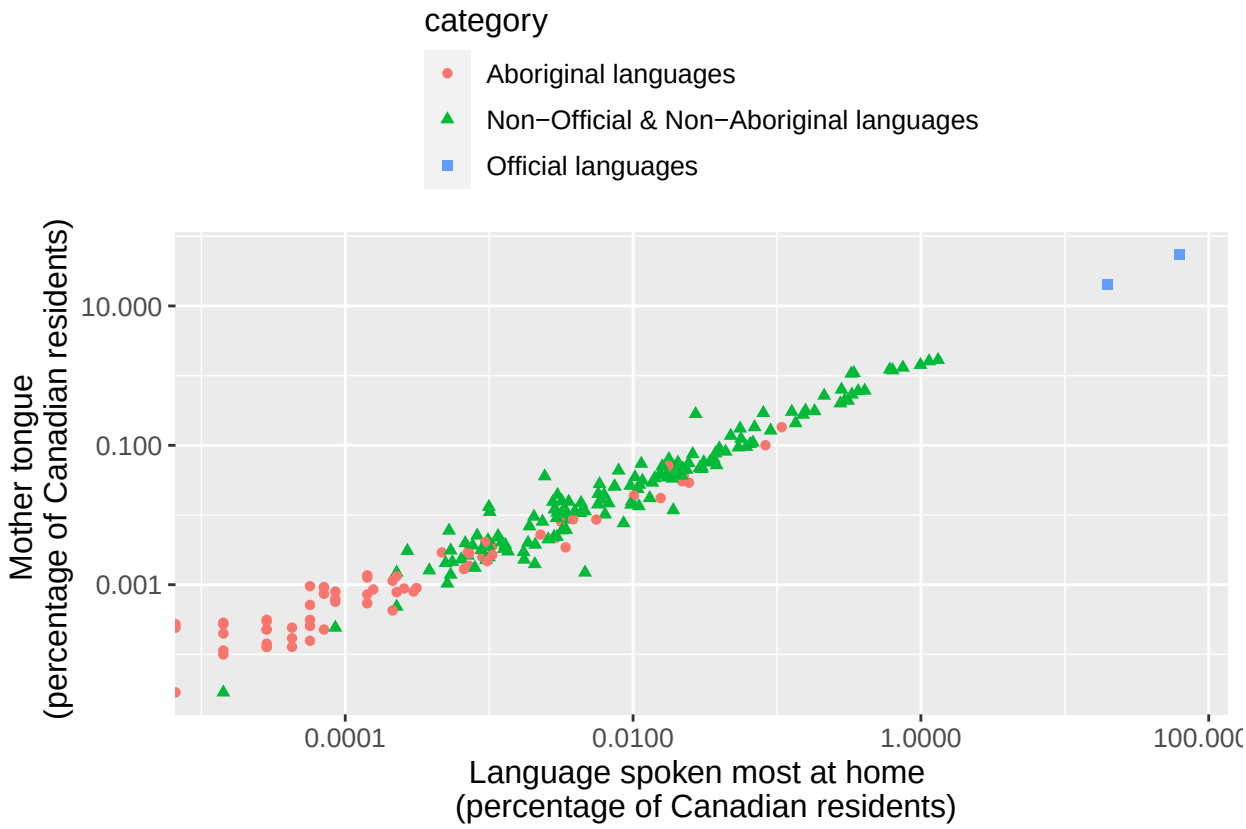
```
## Warning: Transformation introduced infinite values in continuous x-axis
```



Scatterplot with Percents and Colors

```
ggplot(can_lang, aes(x = most_at_home_percent,
                     y = mother_tongue_percent,
                     color = category, shape=category)) +
  geom_point() +
  xlab("Language spoken most at home \n (percentage of Canadian residents)") +
  ylab("Mother tongue \n (percentage of Canadian residents)") +
  theme(text = element_text(size = 12),
        legend.position = "top",
        legend.direction = "vertical") +
  scale_x_log10(labels = comma) +
  scale_y_log10(labels = comma)
```

```
## Warning: Transformation introduced infinite values in continuous x-axis
```


category
Aboriginal languages
Non−Official & Non−Aboriginal languages
Official languages
Mother tongue
(percentage of Canadian residents)
10.000
0.100
0.001
0.0001
0.0100
1.0000
100.000
Language spoken most at home
(percentage of Canadian residents)