First Letters in Pet Names

```
library(tidyverse)
library(dplyr)

# Read and display data
seattle_pets <- readr::read_csv(paste0("https://raw.githubusercontent.com/",
                "rfordatascience/tidytuesday/master/data/",
                "2019/2019-03-26/seattle_pets.csv"))

# Dataset variables
colnames(seattle_pets)
```

```
## [1] "license_issue_date" "license_number"     "animals_name"
## [4] "species"            "primary_breed"      "secondary_breed"
## [7] "zip_code"
```

```
head(seattle_pets[, 1:length(seattle_pets)])
```

```
## # A tibble: 6 x 7
##   license_issue_d~ license_number animals_name species primary_breed
##   <chr>            <chr>          <chr>        <chr>   <chr>
## 1 November 16 2018 8002756        Wall-E       Dog     Mixed Breed,~
## 2 November 11 2018 S124529        Andre        Dog     Terrier, Jac~
## 3 November 21 2018 903793         Mac          Dog     Retriever, L~
## 4 November 23 2018 824666         Melb         Cat     Domestic Sho~
## 5 December 30 2018 S119138        Gingersnap   Cat     Domestic Sho~
## 6 December 16 2018 S138529        Cody         Dog     Retriever, L~
## # ... with 2 more variables: secondary_breed <chr>, zip_code <chr>
```

Add a column to data table corresponding to (capitalized) first letter in names, drop any NULL values, and drop names which occur fewer than 3 times (this gets rid of all names not beginning with a letter of the alphabet):

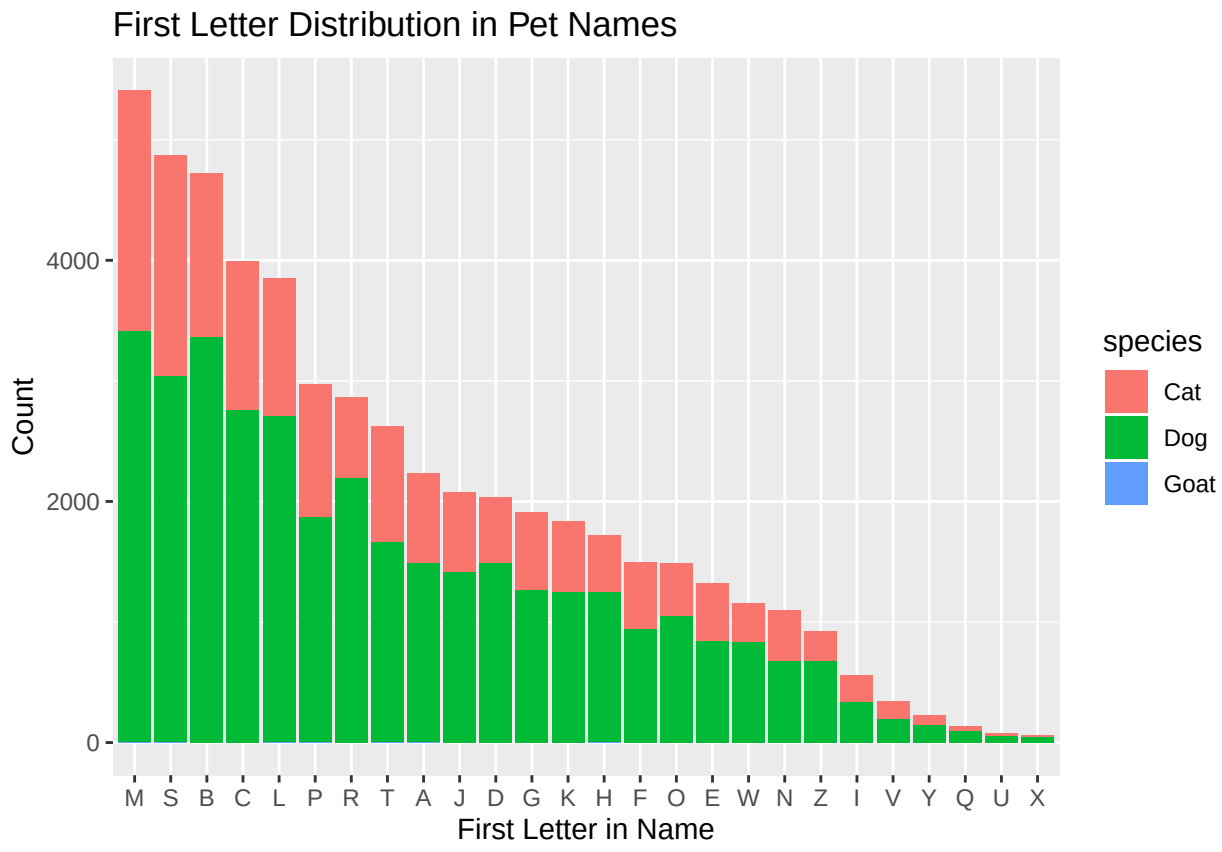
```
seattle_pets <- seattle_pets %>%
  mutate(first_letter = str_sub(animals_name, start = 1, end = 1) %>%
           toupper())

pets_grouped <- seattle_pets %>%
  group_by(first_letter, species) %>%
  count()

pets_grouped <- pets_grouped %>%
  filter(n > 2 & !is.na(first_letter))

head(pets_grouped)
```

```
## # A tibble: 6 x 3
## # Groups:   first_letter, species [6]
##   first_letter species     n
```

```
##   <chr>        <chr>   <int>
## 1 A            Cat       743
## 2 A            Dog      1490
## 3 A            Goat        3
## 4 B            Cat      1356
## 5 B            Dog      3366
## 6 C            Cat      1234
```

Plot distribution of first letter in names

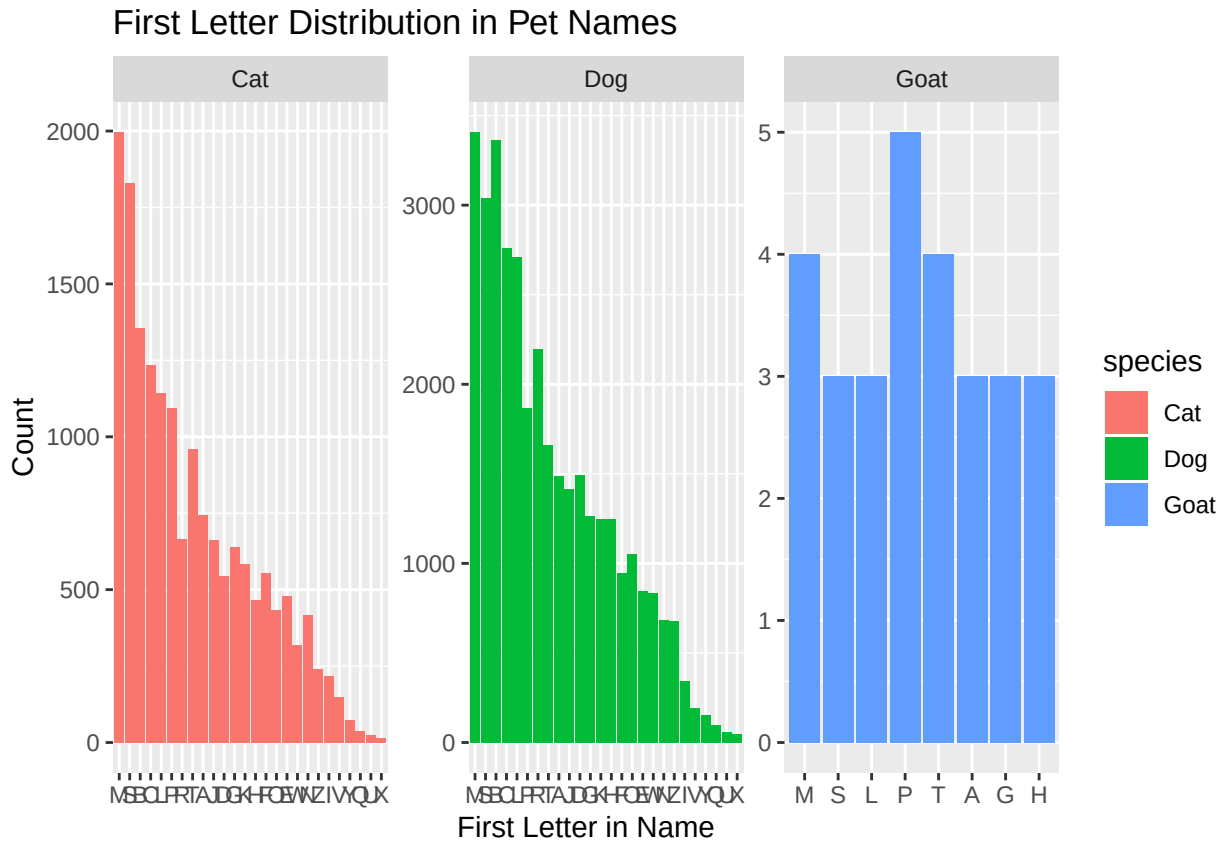
```
pets_grouped_plot <- pets_grouped %>%
  ggplot(aes(reorder(first_letter, -n, FUN = sum), n, fill = species)) +
  geom_col() +
  labs(x = "First Letter in Name", y = "Count") +
  ggtitle("First Letter Distribution in Pet Names")

print(pets_grouped_plot)
```




```
ggsave("firstLetterDistr.png")
```

Looking at each species separately with fancy facet plots:

```
pets_grouped_plot +
  facet_wrap(~pets_grouped$species, scales = "free")
```




```
ggsave("firstLetterFacets.png")
```