Visualizing Distributions and Categorical Data With ggplot2

New Functions covered:
geom_bar()
coord_flip()
scale_y_continuous()
facet_grid()
geom_jitter()
geom_smooth()
filter() with %in% and logical operators

Load tidyverse (as usual)
library(tidyverse)

####

Quick look at the data with glimpse()
glimpse(gss_cat)

Check the first few rows of the dataset
head(gss_cat)

Check the *last* few rows of the dataset
tail(gss_cat)

Make a barplot of the rincome variable
gss_cat |>
  ggplot(aes(x = rincome)) +
  geom_bar(fill = "sienna2")

Make a barplot of the rincome variable switching axes
gss_cat |>
  ggplot(aes(x = rincome)) +
  geom_bar(fill = "sienna2") +
  coord_flip()

Make a histogram of tvhours
gss_cat |>
  ggplot(aes(x = tvhours)) +
  geom_histogram(bins = 10)

Make a histogram of age
gss_cat |>
  ggplot(aes(x = age)) +
  geom_histogram(bins = 10,
                 fill = "sienna2")

Make categories of age then make a bar plot of categories
gss_cat |>
  mutate(agecat = cut_interval(age, 4)) |>
  ggplot(aes(x = agecat)) +
  geom_bar()

```
## Make the same barplot but with axes switched
gss_cat |>
   mutate(agecat = cut_interval(age, 4)) |>
   ggplot(aes(y = agecat)) +
   geom_bar()

## Make the same barplot but with NAs filtered out
gss_cat |>
   filter(!is.na(age)) |>
   mutate(agecat = cut_interval(age, 4)) |>
   ggplot(aes(y = agecat)) +
   geom_bar()

## Make a barplot of marital status
gss_cat |>
   ggplot(aes(x = marital)) +
   geom_bar()

## Make a barplot of marital status with proportions
gss_cat |>
   ggplot(aes(x = marital)) +
   geom_bar(aes(y = after_stat(prop),
                group = 1))

## Make a barplot of marital status with proportions
gss_cat |>
   ggplot(aes(x = marital)) +
   geom_bar(aes(y = after_stat(prop),
                group = 1))

## Make a barplot of marital status with percentages
gss_cat |>
   ggplot(aes(x = marital)) +
   geom_bar(aes(y = after_stat(prop),
                group = 1)) +
   scale_y_continuous(breaks = c(0, 0.4),
                      labels = c("0%", "40%")) +
   labs(y = "Percentage")

## Make a barplot of marital status with percentages using 'scales' package
gss_cat |>
   ggplot(aes(x = marital)) +
   geom_bar(aes(y = after_stat(prop),
                group = 1)) +
   scale_y_continuous(labels = scales::percent_format()) +
   labs(y = "Percentage")


## Look at the distribution of TV hours watched
gss_cat |>
   ggplot(aes(x = tvhours)) +
   geom_histogram(bins = 8)
```

```
## Filter dataset to "Never married" values of marital and then make a histogram
## of tvhours
gss_cat |>
  filter(marital == "Never married") |>
  ggplot(aes(x = tvhours)) +
  geom_histogram(bins = 8)


## Filter dataset to "Married" values of marital and then make a histogram of
## tvhours
gss_cat |>
  filter(marital == "Married") |>
  ggplot(aes(x = tvhours)) +
  geom_histogram(bins = 8)

## Filter dataset to Separated, Divorced, Widowed and make a histogram of
## tvhours
gss_cat |>
  filter(marital %in% c("Separated", "Divorced", "Widowed")) |>
  sample_n(10) ## Check first to see our filtering is working correctly

gss_cat |>
  filter(marital %in% c("Separated", "Divorced", "Widowed")) |>
  ggplot(aes(x = tvhours)) +
  geom_histogram(bins = 8)


## Make a histogram facet plot of tvhours by marital
gss_cat |>
  ggplot(aes(x = tvhours)) +
  geom_histogram(aes(y = after_stat(density)),
                 bins = 8) +
  facet_wrap(vars(marital))

## Make a boxplot of tvhours by marital
gss_cat |>
  filter(marital != "No answer") |>
  ggplot(aes(x = marital, y = tvhours)) +
  geom_boxplot()

## Make a scatterplot of tvhours vs age
gss_cat |>
  ggplot(aes(x = age, y = tvhours)) +
  geom_point()

## Make a scatterplot of tvhours vs age with alpha
gss_cat |>
  ggplot(aes(x = age, y = tvhours)) +
  geom_point(alpha = 1/10)

## Make a scatterplot of tvhours vs age with jitter
gss_cat |>
  ggplot(aes(x = age, y = tvhours)) +
```

```
    geom_jitter()

## Combine ages into categories and make boxplots of tvhours by agecat
gss_cat |>
    filter(!is.na(age)) |>
    mutate(agecat = cut_interval(age, 10)) |>
    ggplot(aes(x = agecat, y = tvhours)) +
    geom_boxplot()

## Make a scatterplot of tvhours vs age with jitter and smooth
gss_cat |>
    ggplot(aes(x = age, y = tvhours)) +
    geom_jitter() +
    geom_smooth()

## Make a smooth of tvhours vs age (no points)
gss_cat |>
    ggplot(aes(x = age, y = tvhours)) +
    geom_smooth()
```