Marriage & Weather datasets advanced ggplot visuals

Luke Geel

10/11/2020

```r
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.0
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(nycflights13)
```

Question 1 A. Use a bar chart to show the distribution of bg_married within each treatment condition. Which treatment condition has the highest proportion of married people?

According to the bar chart, the treatment condition which has the highest proportion of married people is lockbox.

B. Add appropriate title, labels of axes, lengend name.

C. Change the names of the two levels in the legend.

D. Make the following plot.

```r
rosca <- read.csv("/Users/lukegeel/Downloads/rosca (1).csv")

rosca$treatment <- NA
rosca$treatment[rosca$encouragement == 1] <- "control"
rosca$treatment[rosca$safe_box == 1] <- "safebox"
rosca$treatment[rosca$locked_box == 1] <- "lockbox"
rosca$treatment <- as.factor(rosca$treatment)
rosca <-  filter(rosca, has_followup2 == 1 | has_followup2 == "has_followup2")
rosca$bg_female <-  as.factor(rosca$bg_female)
rosca$bg_married <-  as.factor(rosca$bg_married)
```

```r
levels(rosca$treatment)
```

```
## [1] "control" "lockbox" "safebox"
```

```r
ggplot(filter(rosca, bg_married ==1))+
  geom_bar(mapping = aes(x = bg_married, fill = treatment))+
  labs(title = "Marriage vs Treatment", x = "Married people", y = "Count", fill = "Treatment type")
```

Marriage vs Treatment

Count

300

200

100

0

1

Married people

Treatment type

control
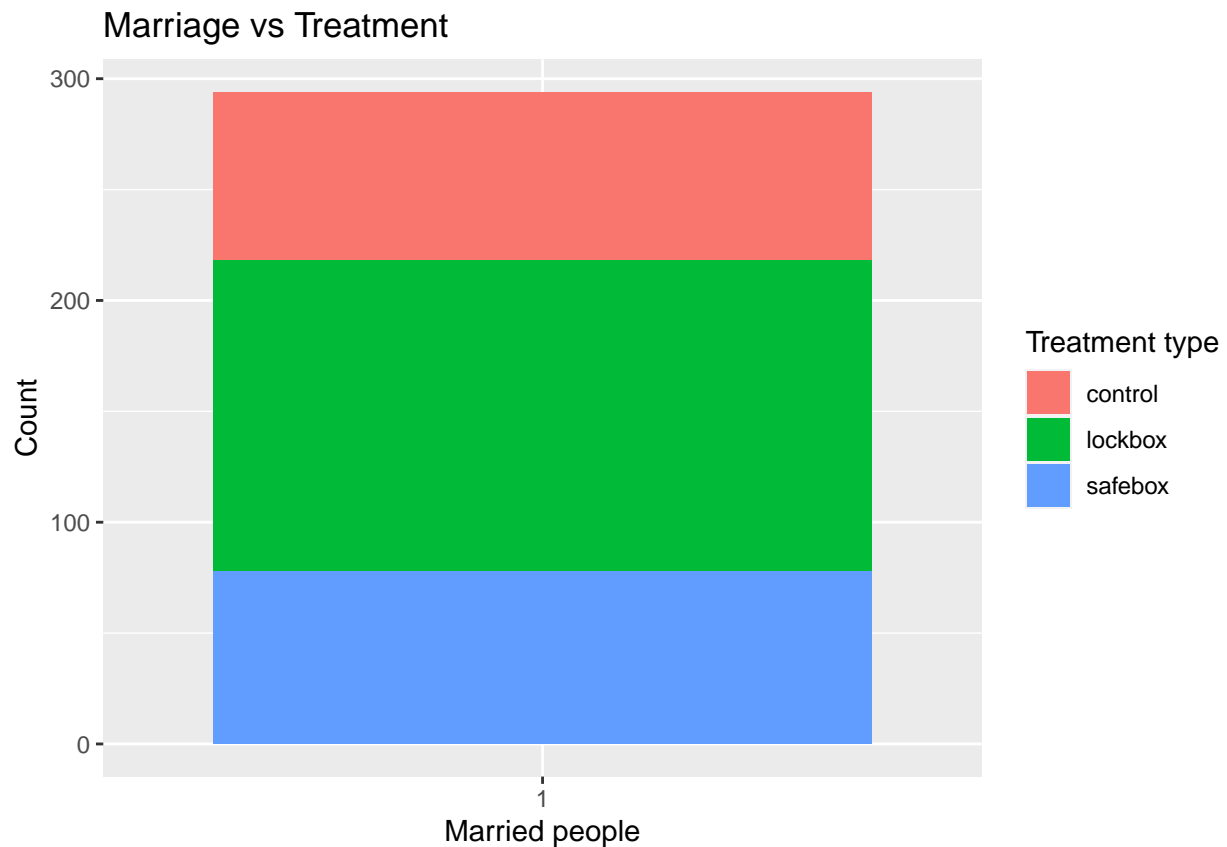
lockbox

safebox

```
ggplot(filter(rosca, treatment == "control" | treatment == "safebox" | treatment == "lockbox"))+
  geom_bar(mapping = aes(x = treatment, fill = bg_married), position = "dodge")+
  labs(title = "Distribution of marital status",
       fill = "marital status", y="", x="")+
  theme(legend.position="bottom")+
  scale_fill_manual(name = "marital status",
                    labels = c("unmarried", "married"),
                    values =c("0"="red","1"="blue")
                    )
```

Distribution of marital status


marital status ■ unmarried ■ married
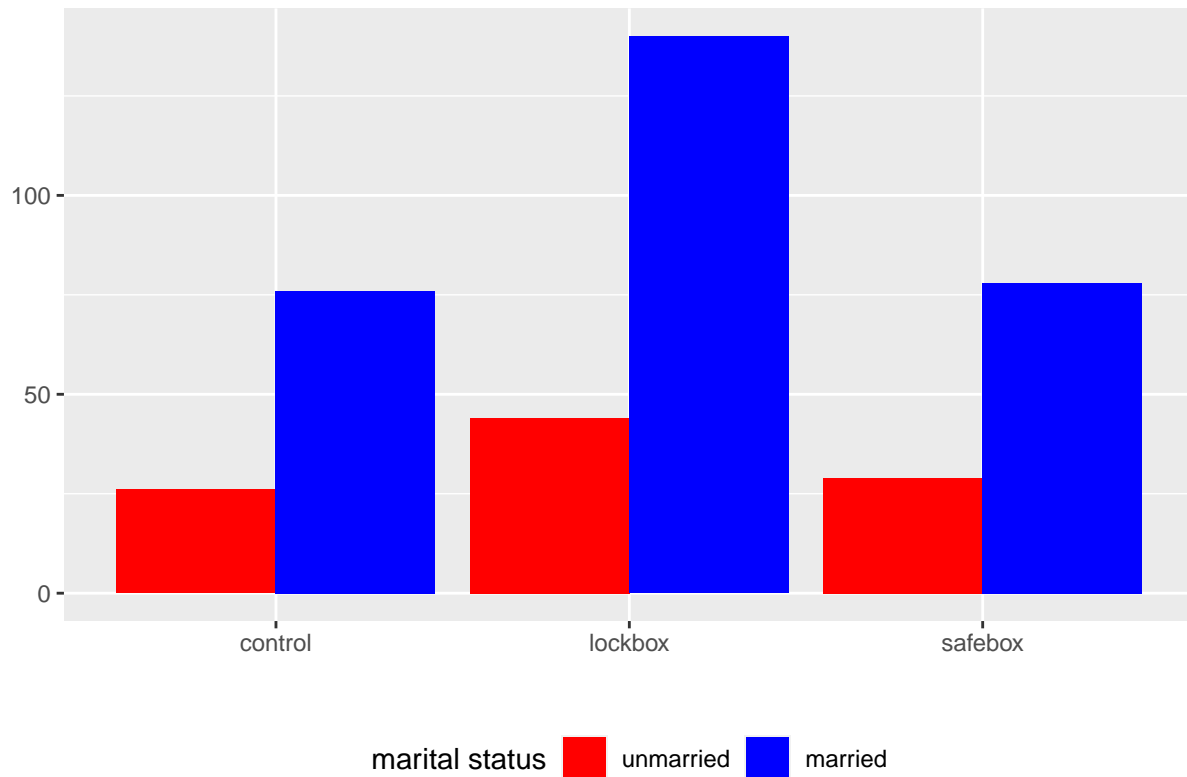
Question 2 A. Add appropriate title, labels of axes. Remove the y-axis ticks and tick labels (200,400...) B. Change the facet labels from numeric month to month abbreviations,e.g., 1 to Jan. C. Make the following plot. You can use round(tapply(weather$temp$, factor$(weather$month),mean)) to obtain the average temperature of each month.

```r
weather <- weather %>% filter(!is.na(temp))
month.labs <- c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12")
names(month.labs) <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sept", "Oct", "Nov", "De
temp.labs <- c("25", "50", "75", "100")
names(temp.labs) <- c("25°", "50°", "75°", "100°")

temp.facet <- ggplot(data = weather)+
  geom_histogram(mapping = aes(x = temp),color= "white",binwidth = 5)+
  facet_wrap(~month,nrow=4)+
  theme(axis.text.y = element_blank(), axis.ticks.y = element_blank())+
  labs(title = "Temperature in 12 months", x = "Temperature (in Fahrenheit)", y="")

temp.facet
```

Temperature in 12 months

Temperature (in Fahrenheit)
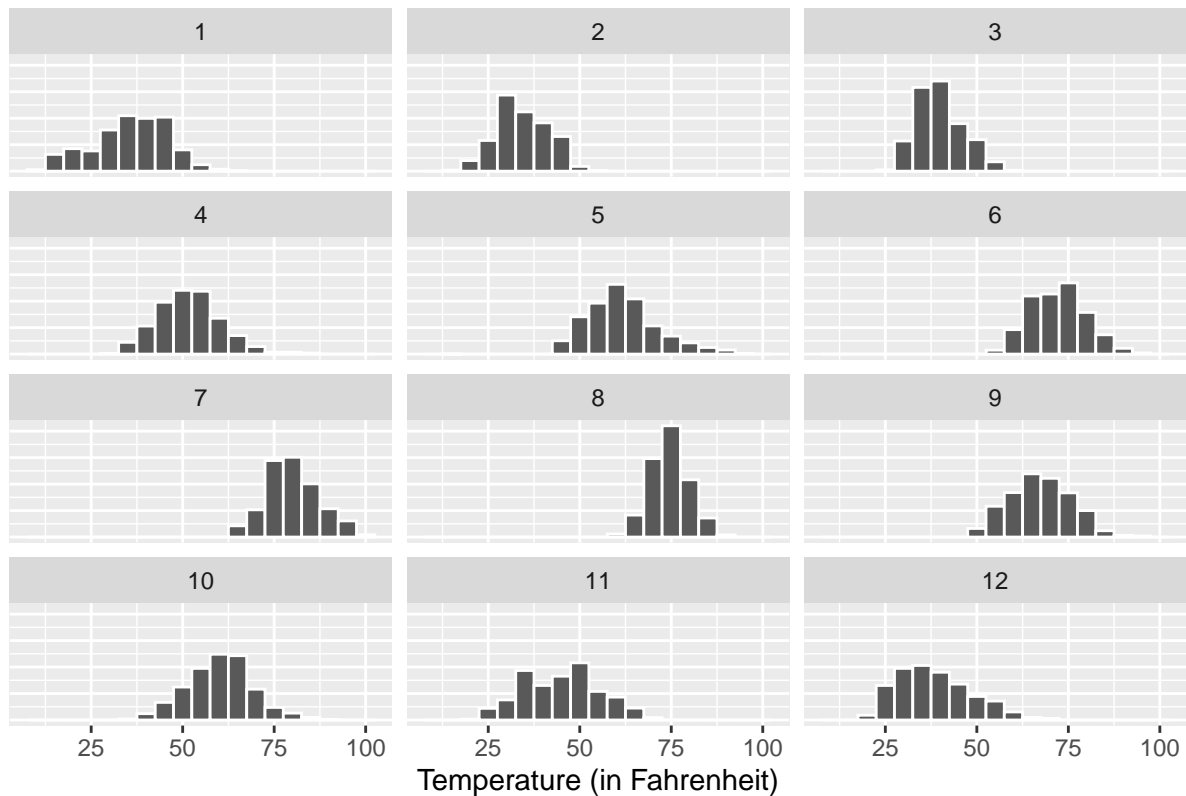
Question 3 A. Remove the x-axis label, and change the tick label from numeric month to month abbreviation as in the following plot. B. Remove the y-axis label, and change the tick label as in the following plot. C. We can make it better by adding the degree symbol. Make the following plot.

```
ggplot(data = weather, mapping = aes(x= factor(month),y = temp)) +
  geom_boxplot()+
  labs(x="",y="")+
  scale_y_continuous(breaks = c(25, 50, 75, 100), labels = c("25°F", "50°F", "75°F", "100°F"))+
  scale_x_discrete(labels = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sept", "Oct", "N
  )
```