STA304A1

Ray

06/05/2021

Abstract

Your Abstract .....

```
library(openintro)
```

```
## Loading required package: airports
## Loading required package: cherryblossom
## Loading required package: usdata
```

```
df<-ncbirths
```

Data

Data

Data

Data

abc

efg

The underlying data set is .... There are in total of 1000 oberservations of pregent mothers, including their income, ..., and number of hospital visits during pregnancy.

Here is the histogram of the interested variable: the number of hospital visits during pregnancy.

```
df %>% ggplot(aes(x = visits)) + geom_histogram(fill = "blue", color = "black", bins = 30) +
  theme_classic()+
  labs(x = "Number of hospital visits during pregnancy", y = "Frequency", title = "Histogram of Number
```

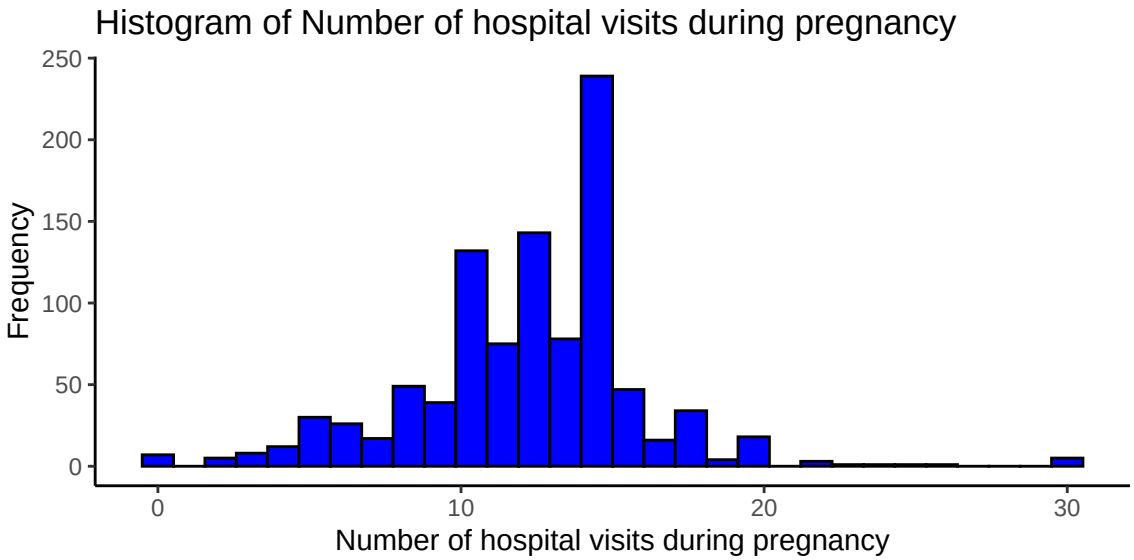


Figure 1: A Plot

STA part: The histogram of the number of hospital visits during pregnancy shows that it is nearly symmetric distributed with some outliers on both (positive and negative) sides of the plot. There is one significant mode at the middle of the plot and one minor mode at the plot's right side. More detail of the number of hospital visits during pregnancy can are summarised below.

```
summary_table <- df %>% filter(!is.na(visits)) %>% summarise(
                                    min = min(visits),
                                    Q1 = quantile(visits,0.25),
                                    median = median(visits),
                                    Q3 = quantile(visits,0.75),
                                    max = max(visits),
                                    IQR = Q3 - Q1,
                                    mean = mean(visits),
                                    sd = sd(visits),
```

Table 1: ABC

min	Q1	median	Q3	max	IQR	mean	sd	Small_Outliers	Large_Outliers
0	10	12	15	30	5	12.10494	3.954934	12	9

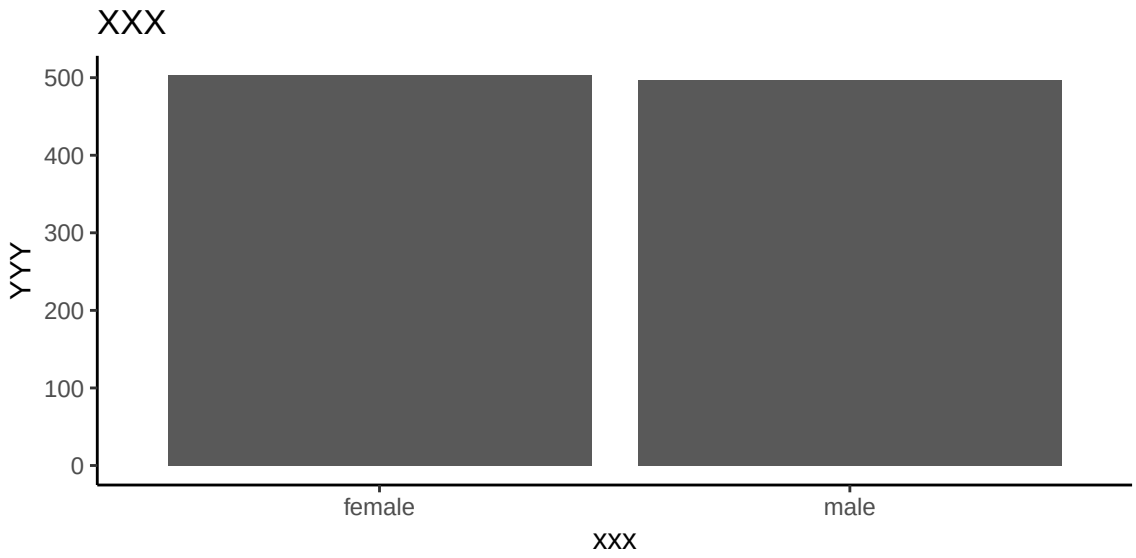
```
                                  Small_Outliers = sum(visits < Q1 - 1.5*IQR),
                                  Large_Outliers = sum(visits > Q3 + 1.5*IQR))
# change their orders, add or remove some of them
kableExtra::kable(summary_table, caption = "ABC")
```

STA part: The mean and median of hospital visits during pregnancy are similar at 12, which supports that the distribution of the number of hospital visits during pregnancy is symmetric. The sample standard deviation (3.9549337) is relatively small, meaning the uncertainty of the data is relatively low. There are 21 outliers in the model.

More non-statistical work. . . .telling a story

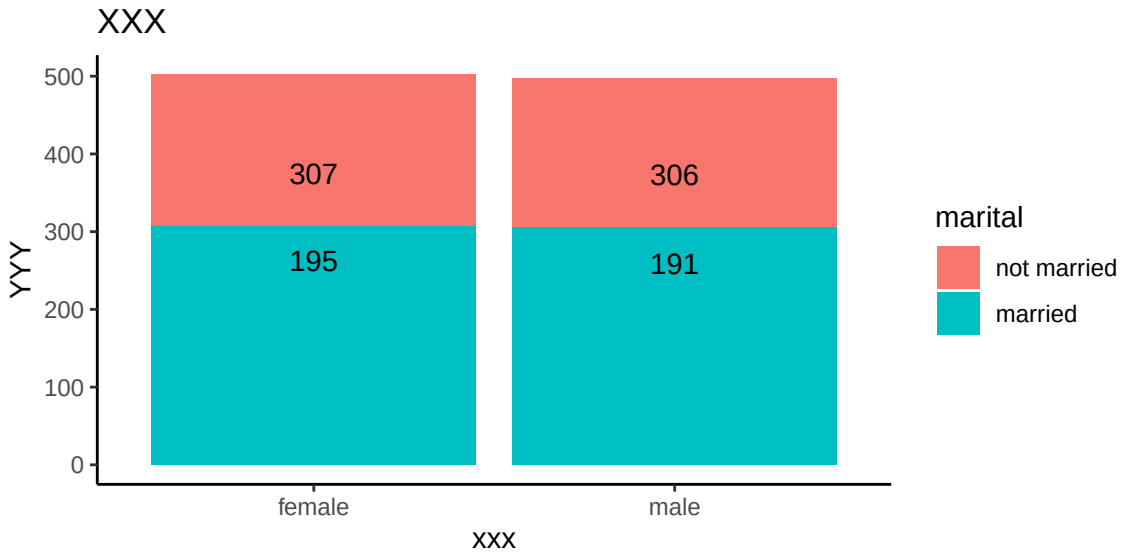
More Plots

```
df %>% ggplot(aes(x= gender)) + geom_bar() + theme_classic()+
  labs(x = "xxx", y = "YYY", title = "XXX")
```



which one is higher...

```
df %>% filter(!is.na(marital)) %>%
  ggplot(aes(x= gender, fill = marital)) + geom_bar() + theme_classic()+
  geom_text(stat = 'count',aes(label =..count.., vjust =-2))+
  labs(x = "xxx", y = "YYY", title = "XXX")
```

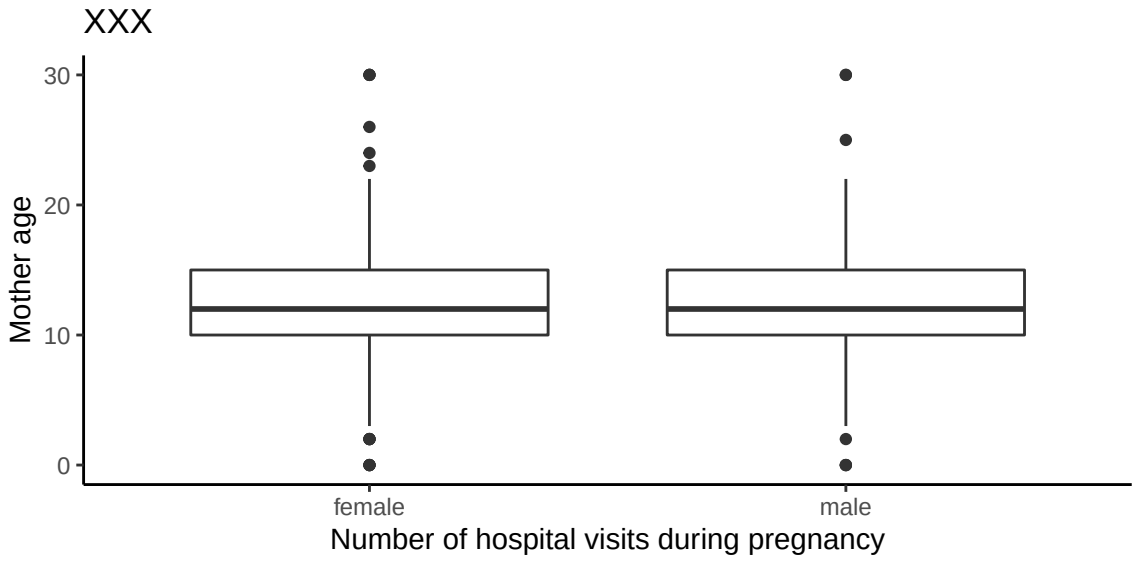


which one is higher...

talk about the proportion

```
df %>% ggplot(aes(y = visits, x= gender)) + geom_boxplot() + theme_classic()+
  labs(x = "Number of hospital visits during pregnancy", y = "Mother age", title = "XXX")
```

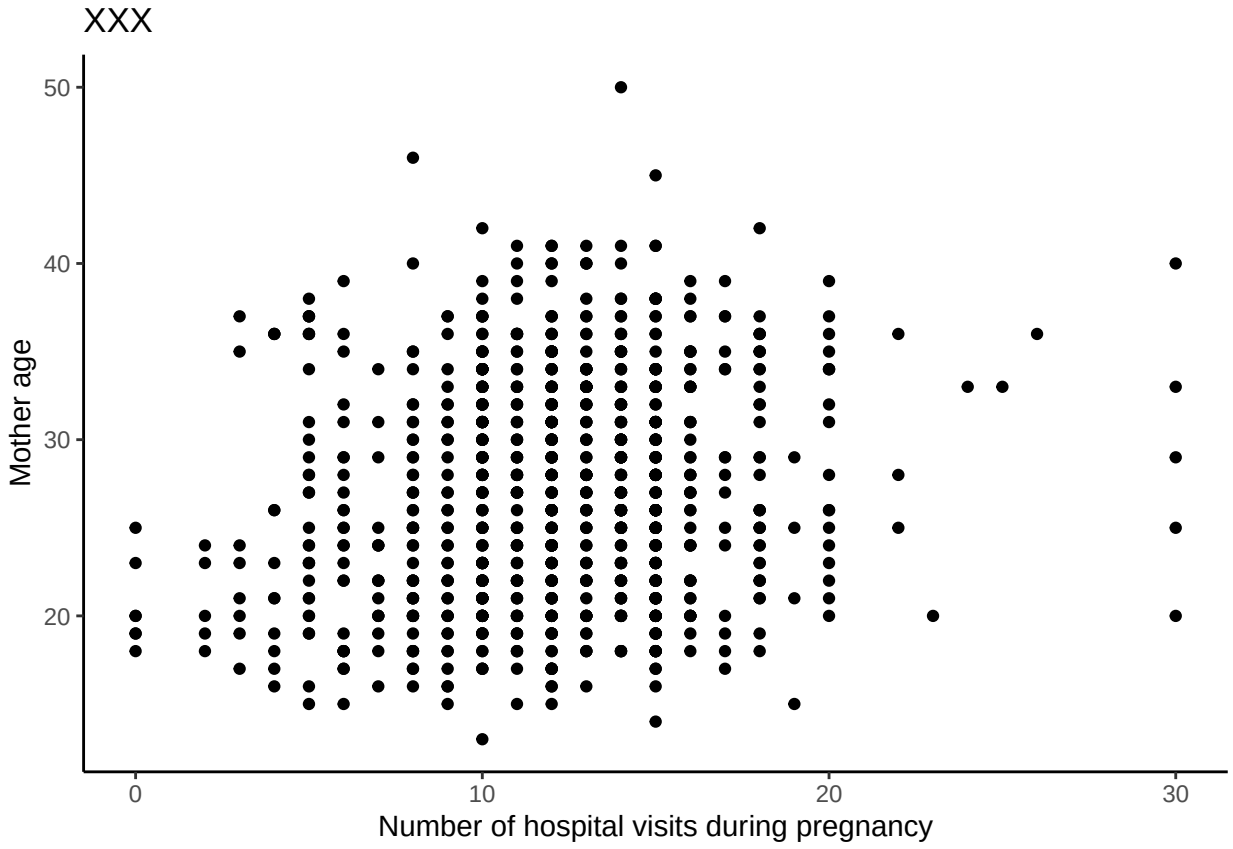
```
## Warning: Removed 9 rows containing non-finite values (stat_boxplot).
```



compare: median, IQR, outliers, range...

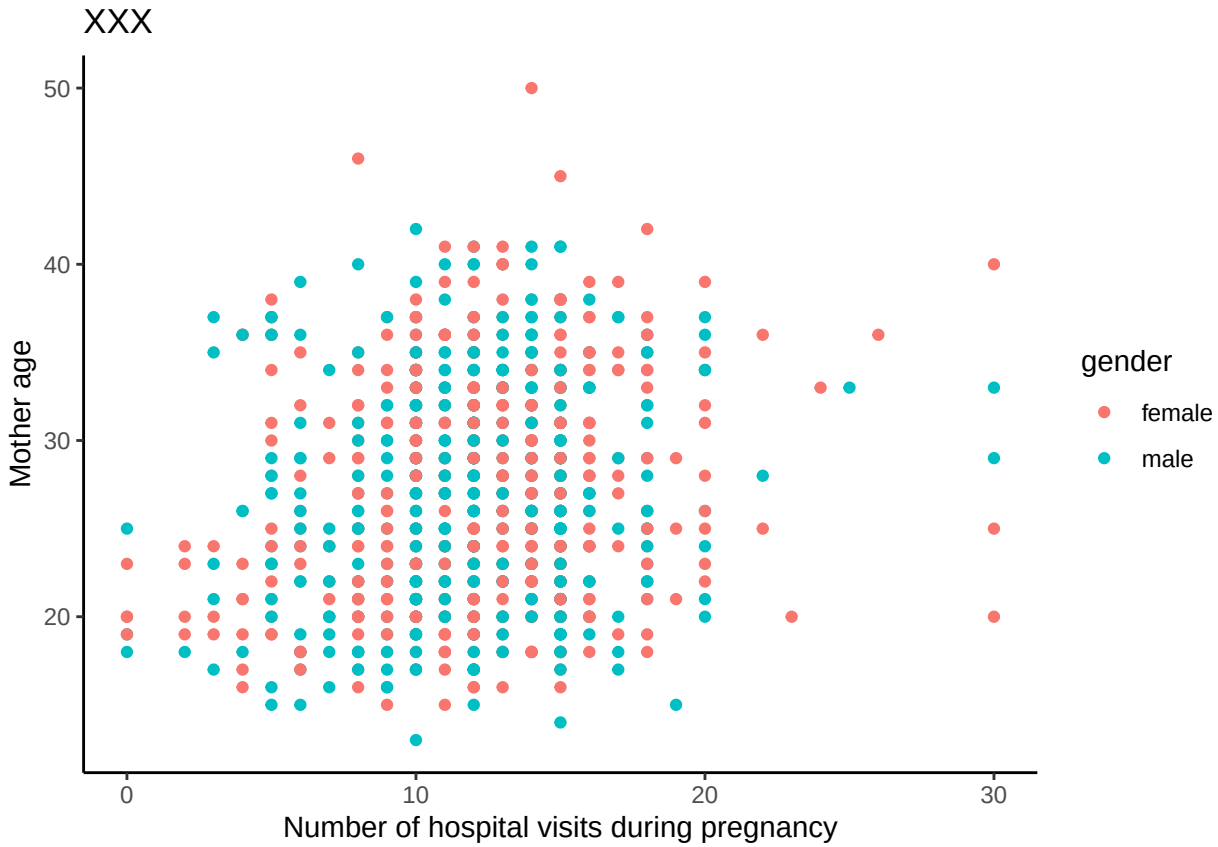
```
df %>% ggplot(aes(x = visits, y = mage)) + geom_point() + theme_classic()+
  labs(x = "Number of hospital visits during pregnancy", y = "Mother age", title = "XXX")
```

```
## Warning: Removed 9 rows containing missing values (geom_point).
```




```
df %>% ggplot(aes(x = visits, y = mage,color = gender)) + geom_point() + theme_classic()+
  labs(x = "Number of hospital visits during pregnancy", y = "Mother age", title = "XXX")
```

```
## Warning: Removed 9 rows containing missing values (geom_point).
```



talk about direction, form, strength

- Need other plots? draw it by hand first, then google it or ask me for the code
- Remember to tell a story! not just statistics!

More tables

```
summary_table <- df %>% filter(!is.na(visits)) %>%
  group_by(gender) %>% summarise(min = min(visits),
                                 Q1 = quantile(visits,0.25),
                                 median = median(visits),
                                 Q3 = quantile(visits,0.75),
                                 max = max(visits),
                                 IQR = Q3 - Q1,
                                 mean = mean(visits),
                                 sd = sd(visits),
                                 Small_Outliers = sum(visits < Q1 - 1.5*IQR),
                                 Large_Outliers = sum(visits > Q3 + 1.5*IQR))
# change their orders, add or remove some of them
knitr::kable(summary_table)
```

gender	min	Q1	median	Q3	max	IQR	mean	sd	Small_Outliers	Large_Outliers
female	0	10	12	15	30	5	12.26358	4.131037	8	6
male	0	10	12	15	30	5	11.94534	3.766922	4	3

Tables	Are	Cool
col 3 is	right-aligned	$1600
col 2 is	centered	$12
zebra stripes	are neat	$1

- more table format see https://rfortherestofus.com/2019/11/how-to-make-beautiful-tables-in-r/

- warning=FALSE, message=FALSE, echo = FALSE
- There should be no raw code. Any output should be nicely formatted.
- Remember to end each section with a concluding sentence. This means reiterating the key points from your writing.

Something Wickham et al. (2019).

(Wickham et al. 2019)

A (2019)

Reference

A, B & C. 2019. “XXX,” 1686.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.