

Homework #2

Francisco Brady

2024-09-10

Exercises

2.10

2.13

2.14

2.17

2.18

2.27

2.28

2.31

Computing Question

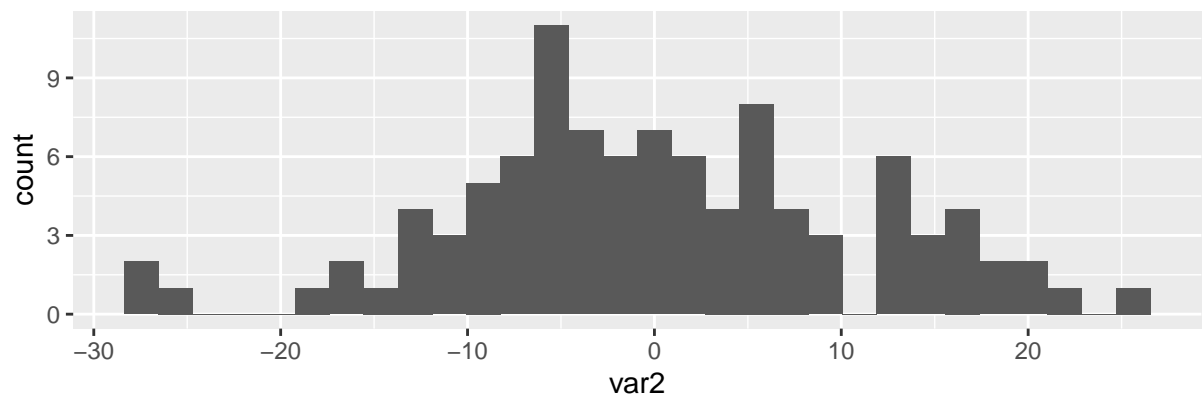
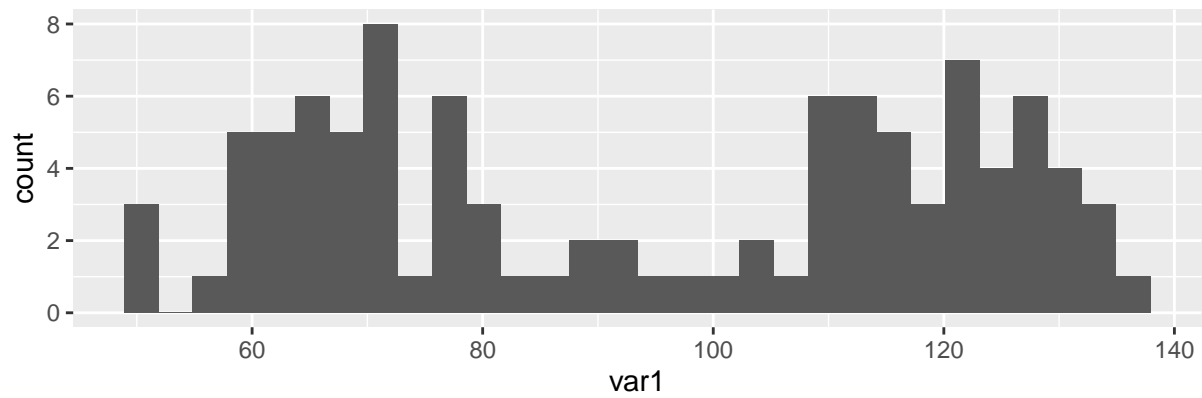
```
data <- read_csv('hw1_data.csv')
```

```
## Rows: 100 Columns: 2
## -- Column specification -----
## Delimiter: ","
## dbl (2): var1, var2
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
var1hist <- ggplot(data) + geom_histogram(aes(var1))
var2hist <- ggplot(data) + geom_histogram(aes(var2))
var1hist / var2hist
```

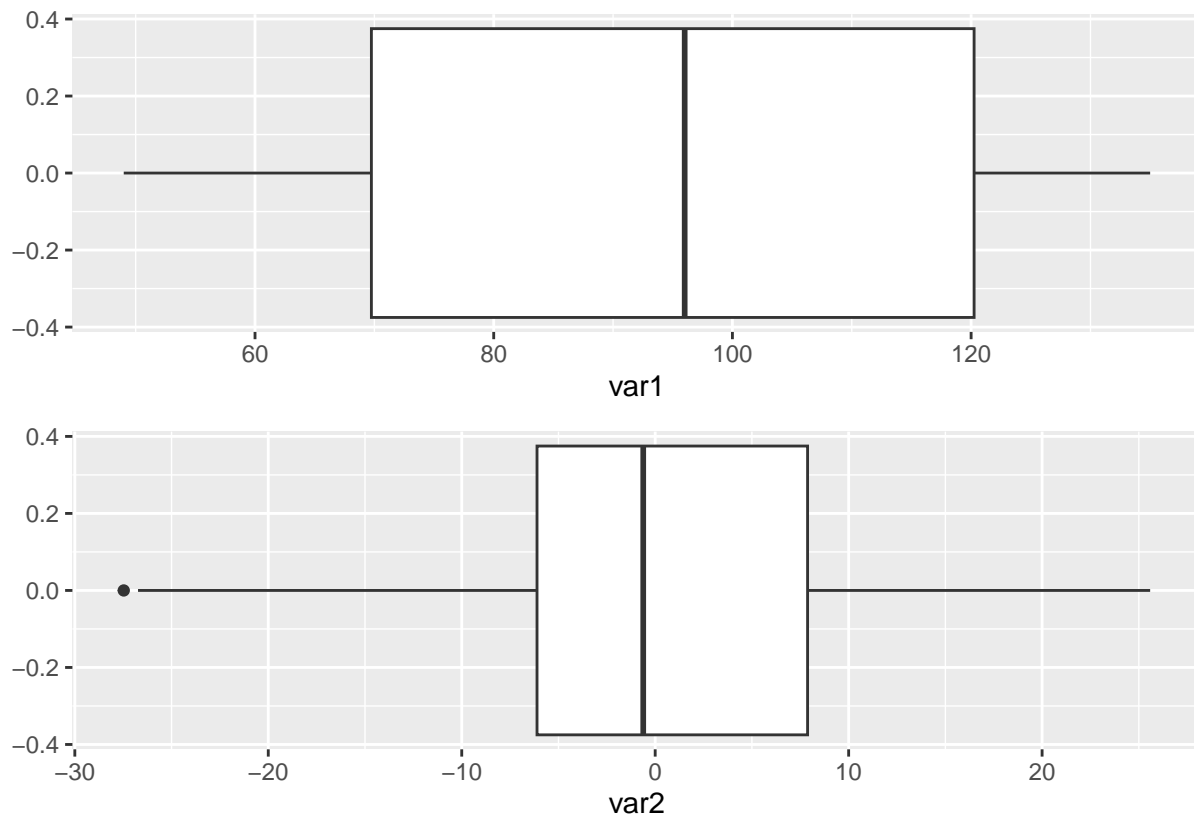
Plot each variable using a histogram and a boxplot. Describe the distributions of var1 and var2.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Var 1 is bimodal in distribution. The values range from about 50 to around 140.
 Var 2 has a more unimodal distribution. The range is from -30 to a little bit over 20.

```
var1box <- ggplot(data) + geom_boxplot(aes(var1))
var2box <- ggplot(data) + geom_boxplot(aes(var2))
var1box / var2box
```



Var 1 has an IQR range of 120 to around 70 (50 units). The median of var 1 is a little over 95. Var 2 has a smaller IQR and a tighter distribution. The median of var 2 looks to be around -1.

What type of plot do you think best illustrates each distribution? Why? The histogram is more useful for showing the distribution of the data in var 1. Relying only on the boxplot would obscure this about var 1. Because var 2 has a more unitary distribution and longer tails, the boxplot is appropriate to illustrate the spread of the data.

```
bind_rows(summary(data$var1),
           summary(data$var2))
) %>% knitr::kable(row.names = TRUE)
```

Compute descriptive statistics for each variable. Use what you know about the distributions of the variables to decide which descriptive statistics are most appropriate.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1	49.00000	69.75000	96.00000	94.69000	120.25000	135.00000
2	-27.47148	-6.10797	-0.61728	0.30899	7.87804	25.58492

The distribution of Var 1 is bimodal. The mean is less than the median, which indicates a slight left skew,