****— title: “Lab 5” author: “” date: “2023-06-14” output: html_document —

Question 1

Load the data wk5.csv into R and save it as d5.

```
#(a)

library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.2     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.2     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
#(b) Input data file
d5 <- read.csv('data_wk5.csv')
```

```
#(c) Determine the minimum and maximum values of d5. You do not need to save these but
#    you must show that you printed them out.
quantile(d5$x)
```

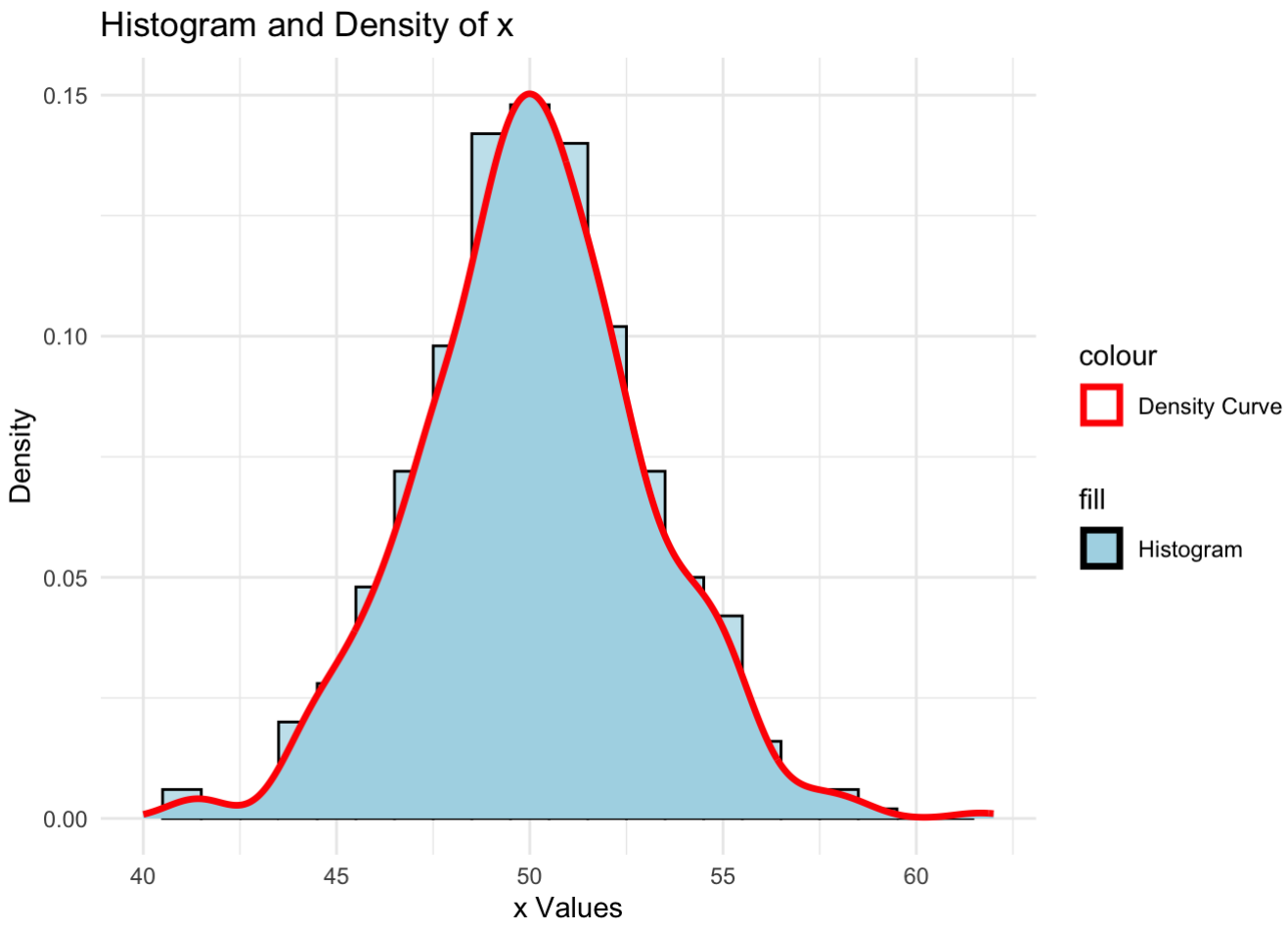
```
##       0%      25%      50%      75%     100%
## 40.88721 48.27510 50.09486 51.90116 61.65908
```

```
#(d) Create a ggplot2 histogram of the \textit{d5} variable. Give the histogram border and fill colours of your l
iking. Name the $x$-axis ``Values'', and make the title ``Histogram of Values.``


#    Set the minimum and maximum values of the x-axis so that all data points are represented. For example, if th
e minimum and maximum values of the variables are 40.4 and 55.6 respectively, you should set the histogram values
to be 40 and 56 respectively. Set the binwidth to 2.

library(ggplot2)

# Calculate the minimum and maximum values of the x-axis
x_min <- floor(min(d5$x))
x_max <- ceiling(max(d5$x))


# Create the histogram
ggplot(d5, aes(x = x, y = ..density.., fill = "Histogram")) +
  geom_histogram(color = "black", fill = "lightblue", alpha = 0.7, binwidth = 1) +
  geom_density(aes(y = ..density.., color = "Density Curve"), size = 1.2) +
  scale_fill_manual(values = "lightblue") +
  scale_color_manual(values = "red") +
  labs(x = "x Values", y = "Density", title = "Histogram and Density of x") +
  xlim(x_min, x_max) +
  theme_minimal()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## ℹ Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: Removed 2 rows containing missing values (`geom_bar()`).
```



Question 2

This question involves gauging whether or not a distribution seems normally distributed.

a. Does the histogram created in Question 1 seem normally distributed?

The histogram in question one is not normally distributed because it depicts a Leptokurtic distribution. These types of distributions have a ‘kurtosis’ greater than three. The Kurtosis is the ‘peak’ of the mountain also referred to as its tail.

d. After comparing the answers from B and C I am more inclined to say that x is evenly distributed because the the results from both B and C are really close to each other.

```
#(a) Does the histogram created in Question 1 seem normally distributed?

#(b) Find the empirical 2.5% and 97.5% quantiles and save them as qlow and qhigh, respectively.
qlow <- quantile(d5$x, probs = 0.025, na.rm = FALSE)
qhigh <- quantile(d5$x, probs = 0.975, na.rm = FALSE)

qlow
```

```
##     2.5%
## 44.37995
```

```
qhigh
```

```
##    97.5%
## 55.59352
```

```
#(c) Finding the normal theoretical 2.5% and 97.5% quantiles of x.
x_mean <- mean(d5$x)
x_sd <- sd(d5$x)

qnorm(.025,mean = x_mean, sd = x_sd)
```

```
## [1] 44.36197
```

```
qnorm(.975,mean = x_mean, sd = x_sd)
```

```
## [1] 55.83004
```

```
#(d) Are your answers in (b) and (c) close?
```