Math 338 **Lab Assignment #23** Fall 2019

In this lab, we will graphically assess the bone strength (in $cm^4/1000$) for a random sample of young men. Specifically, the bone strength of each person's dominant and non-dominant arm was recorded for two groups of young men: those who do not play baseball (Control Group) and those who do (Baseball Group) – see Exercises 2.18 and 2.19 in your book. The data can be found on Titanium as the file *armstrength.txt*.

Import the armstrength.txt file into RStudio. In the `Import` dialog, change the `Delimiter` to Tab.

Let's start by making a scatter plot of the data with Nondominant as the x and Dominant as the y. As usual, replace the question marks in order to complete the plot.

```
> library(ggplot2)
```

```
> arm_plot <- ggplot(data = armstrength, mapping = aes(x = Nondominant, y =
Dominant)) + geom_point() + labs(title = "?", x = "?", y = "?")
```

```
> print(arm_plot)
```

Question #1 Insert the plot below.

Please see first attached plot after lab questions.

Question #2 Describe the form, direction and strength of the relationship. Use more technical terms where appropriate, and reasonable adjectives otherwise.

The form of the graph is linear and positively correlated. The strength of the graph is strong towards the base but as it rises the data becomes dispersed, having a weaker strength.

Question #3 Just looking at the plot, do you see any outliers? If so, describe where they are on the plot.

There does not seem to be any apparent outliers by looking at the data.

Next, let's obtain the equation of the regression line. The `lm()` command creates the model:

```
> model1 <- lm(Dominant ~ Nondominant, data = armstrength)
```

And the `coef()` command gets the coefficients out:

```
> coefs <- coef(model1)
```

We can get a lot more information using the `summary()` command:

```
> summary(model1)
```

Question #4 Copy and paste the `Coefficients:` table from the summary below.

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	**-1.853**	**3.601**	**-0.515**	**0.611**	
Nondominant	**1.383**	**0.205**	**6.745**	**2.53e-07**	***

Question #5 Identify the values of the slope and intercept from the output in **Question #4**, and using those values, write out the least squares regression equation for predicting y. Use the convention of replacing x and $\hat{y}$ with the actual variable names/descriptors.

Slope: 1.382817

Y-Intercept: -1.853414

y = 1.382817x - 1.853414

Question #6 What are the predicted values for the dominant arm bone strength for the following corresponding bone strengths for the nondominant arm: 16, 19, 21, and 25 $cm^4/1000$? (Hint: use your regression equation)

y = 1.382817(16) – 1.853414 = 20.27168
y = 1.382817(19) – 1.853414 = 24.420109
y = 1.382817(21)- 1.853414 = 27.185743
y = 1.382817(25) – 1.853414 = 32.717011

Let's overlay the least-squares regression line on a scatterplot.

```
> arm_plot2 <- arm_plot + geom_abline(aes(intercept = coefs[1], slope =
coefs[2]))  # note: coefs is the variable we stored the slope/intercept in

> print(arm_plot2)
```

Question #7 Insert the plot with data and fitted least-squares regression line below.

Please see second attached plot after lab questions.

It makes sense that baseball players might have a different relationship between dominant and non-dominant arm bone strength than non-baseball players. Copy the code you used to create your original scatterplot for **Question #1** and modify it by adding `color` and `shape` arguments as shown below. Change the variable name to `arm_plot3` so that we can still access the original `arm_plot` plot:

```
> arm_plot3 <- ggplot(data = armstrength, mapping = aes(x = Nondominant, y
= Dominant, color = Group, shape = Group)) + geom_point() + labs(title =
"?", x = "?", y = "?")
```

Now we'll change the symbol and color of the points. In the code below, replace `color 1` and `color 2` with the names of colors (all lowercase), or you could use a six-digit hex code (i.e., #000000) instead.

Also, replace **number 1** and **number 2** with numbers between 0 and 25. For a list of which numbers correspond to which shapes, see http://www.cookbook-r.com/Graphs/Shapes_and_line_types/.

```
> arm_plot_color <- arm_plot3 + scale_color_manual(values = c("color 1",
"color 2")) + scale_shape_manual(values = c(number 1, number 2))
```

Finally, let's add the least-squares line by group. Note that here the lines will only cover the range of x-values in each group.

```
> arm_plot_group <- arm_plot_color + geom_smooth(method = "lm", se = FALSE)

> print(arm_plot_group)
```

Question #8 Insert the new plot below.

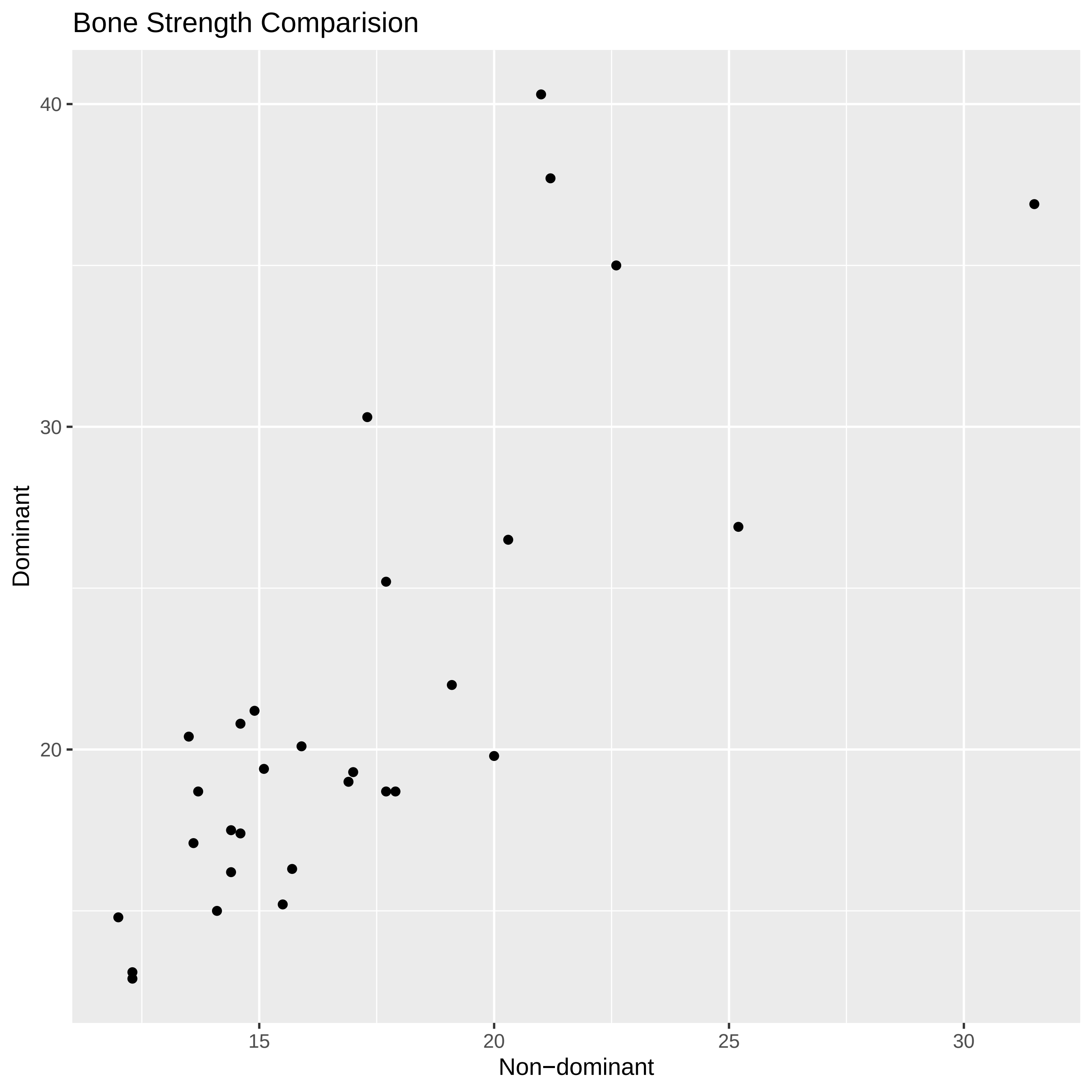
Please see the last page for the plot.

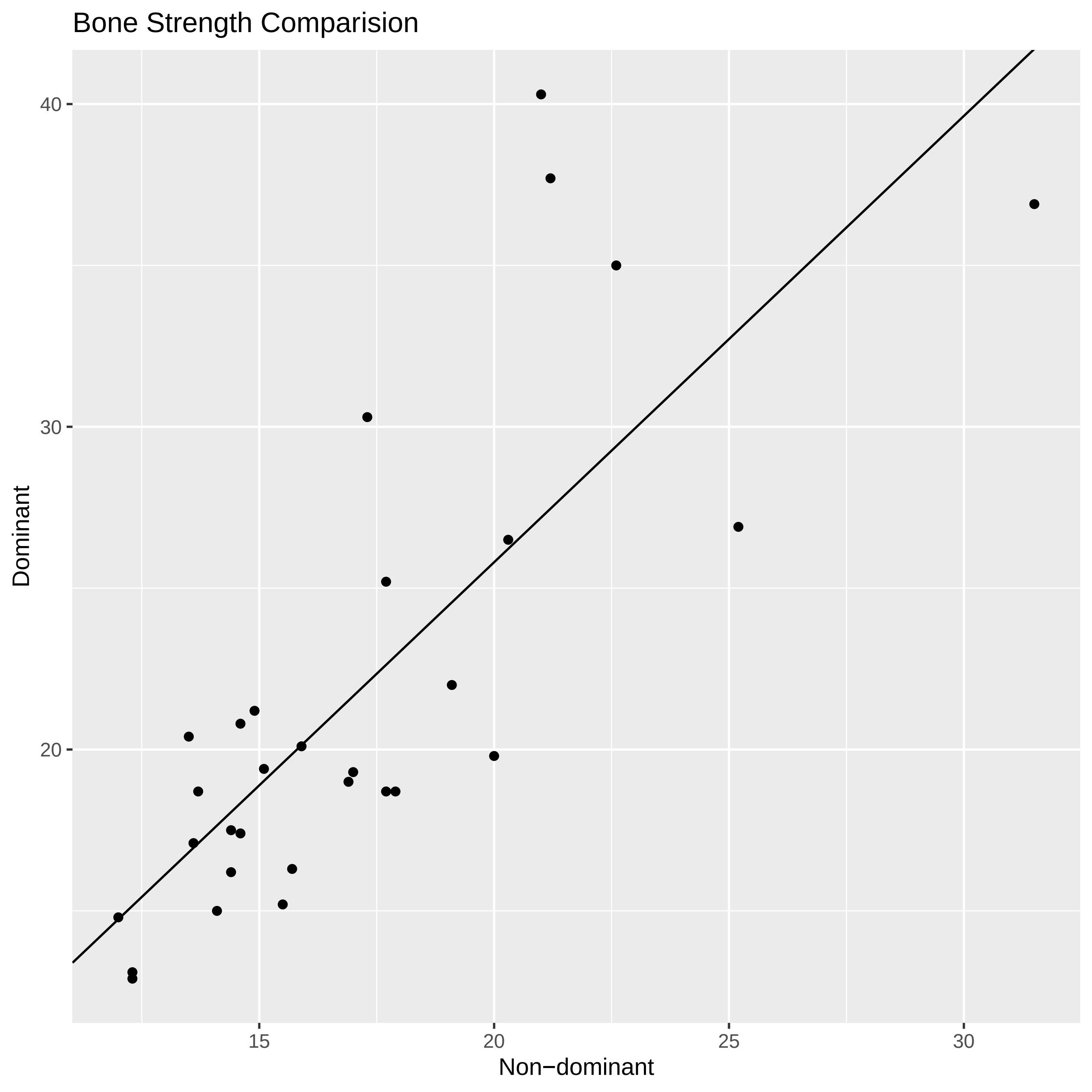
Question #9 Describe any differences between Baseball and Control groups in the direction, form, and strength of the relationship between bone strength of nondominant and dominant arms. Use more technical terms where appropriate, and reasonable adjectives otherwise.

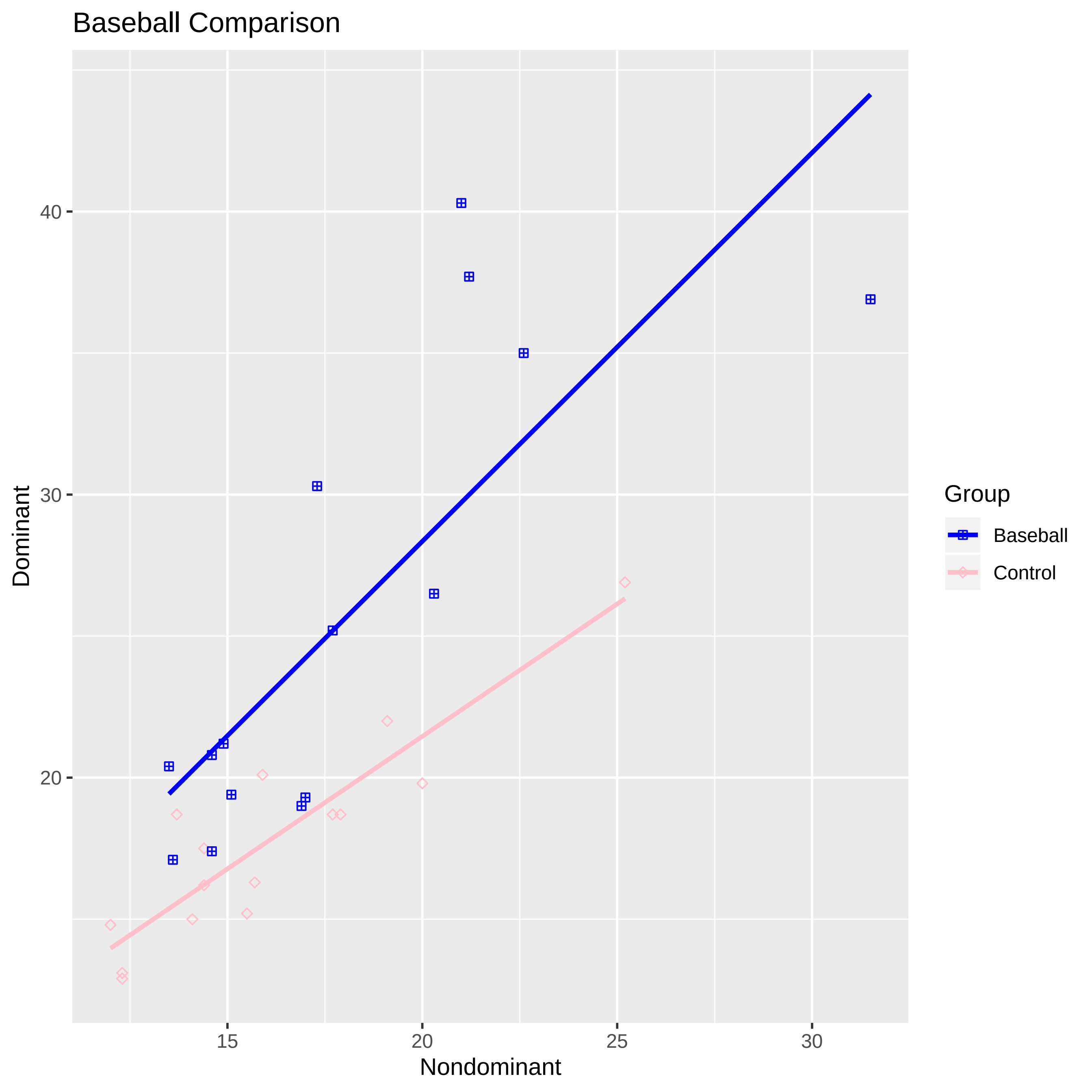
Both the control and baseball groups have strong, positive and linear relationships. However, the dominant and nondominant relationship is stronger in the control group.

Question #10 Write a couple of sentences to summarize what you have learned about Dominant and Nondominant arm strength in Baseball and Control groups. For this answer, try to use as little statistical jargon as you can – explain it like you would to a 10-year-old. Think about how both the values of the individual variables (nondominant and dominant arm bone strength) and the relationship between those two variables might differ between the two groups.

The baseball people have stronger arms because they use their arms more than the people in the control group. Also, this implies that the baseball group have stronger bones. This can be seen in how steep the slope is of the baseball group comparing to the control group slope.


Bone Strength Comparision
Dominant
40
30
20
Non−dominant
15
20
25
30


Bone Strength Comparision
40
30
20
Dominant
15
20
25
30
Non−dominant



Baseball Comparison
Dominant
40
30
20
15
20
25
30
Nondominant
Group
Baseball
Control