Lab 6

David Schultheiss

11-8-2020

Lab 6 will be using the same data as with lab 5. Below, I have included some lines to load in the data and do some cleaning. Your only job for loading in the data is to make sure you have the correct file path (i.e., reference where you have lab 5 data stored), which is denoted by the fill in below.

Question 1

Let's set the stage: Coach Sausaman wants to break down the August 6th practice by drill. He wants to determine the *relative effort* a player devotes to each practice drill with respect to their maximum velocity, total distance, and total player load. This means that we need to standardize the metrics (e.g., percent of total distance that player A spends in 11v11, or warm up, etc. . .), and then divide the standardized metrics by the percent of time a player spent on that drill. However, he want's you to notice that in the data set, the practice drill "11v11" was recorded twice, so you will need to add these two sessions together. As a general outline, you will need to:

1. Filter the rows to select only the one date,
2. Remove the rows that have "Session" in them as they are the total practice summary,
3. Select the appropriate column names,
4. Group by player and practice session,
5. Combine the 11v11 drills,
6. Standardize the metrics,
7. Divide the standardized metrics by time spent,
8. Remove unneeded columns,
9. Change to long format for plotting.

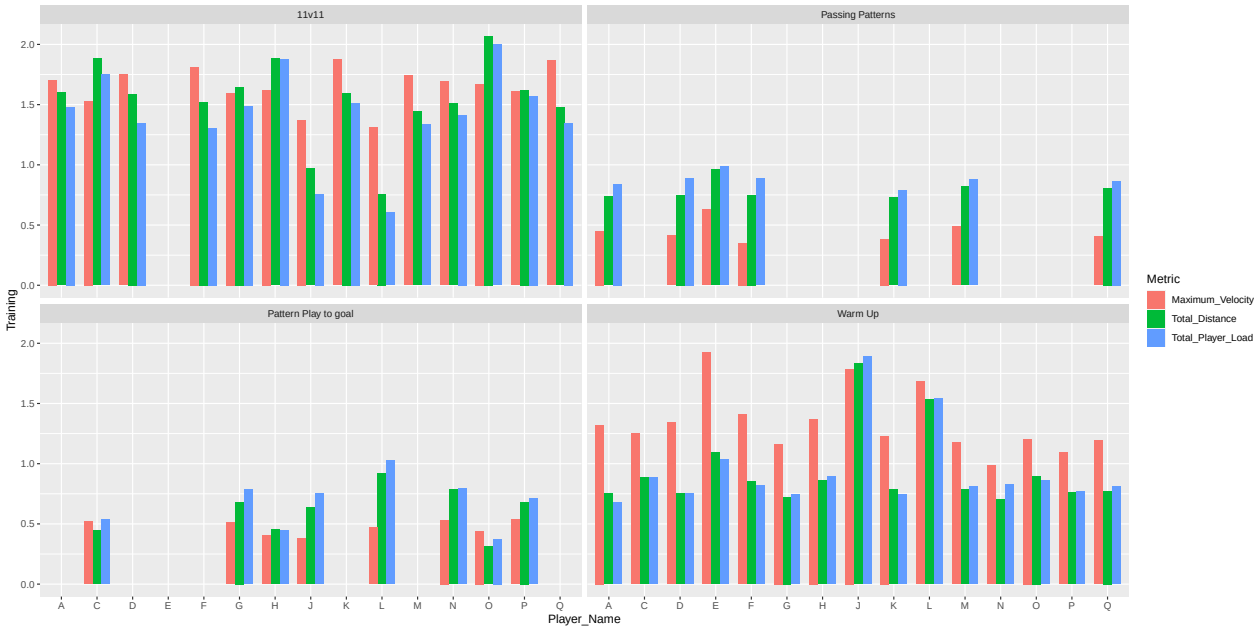
After you have created the data set, you will need to plot it. I have not included any outline for the plot, so it is up to you on how to visualize it. Feel free to make it as basic as you want, but be sure that the plot makes sense and is clean.

Answer

```
ddf =  dat %>%
  filter(Date == '2018-08-06') %>%
  filter(Period_Name != 'Session') %>%
  select(Player_Name, Period_Name, Field_Time, Total_Distance, Maximum_Velocity, Total_Player_Load) %>%
  group_by(Player_Name, Period_Name) %>%
  summarize_at(vars(Field_Time:Total_Player_Load), ~ sum(.)) %>%
  mutate_at(vars(Field_Time:Total_Player_Load), ~ as.numeric(.)/as.numeric(sum(.))) %>%
  ungroup() %>%
  mutate_at(vars(Total_Distance:Total_Player_Load), ~./Field_Time) %>%
  select(-Field_Time) %>%
```

```
  pivot_longer(-c(Player_Name, Period_Name), names_to = 'Metric',
               values_to = 'Training')

ggplot(ddf, aes(x= Player_Name, y= Training, fill= Metric, width= 0.75)) +
  geom_bar(position = 'dodge', stat = 'identity') +
  facet_wrap(~ Period_Name)
```



Question 2

Now, Coach Sausaman wants which practice drills have the highest player load per minute when grouping by position. He also wants to know the 95% confidence intervals to these estimates. He also recognizes that some drills are not as common as others, so he only your plot to show drills that have occurred more than 20 times over the course of the given data set. Breaking this down into bite-sized steps:

1. Select the appropriate columns,
2. Group by practice drill,
3. Keep only the drills that occur 20 or more times,
4. Group by position and drill,
5. Calculate the summary statistics.

Again, you will need to determine how best to visualize the data. However, Coach Sausaman recognizes that the goal keepers have a different practice routine. Therefore, he does not want you to have the goal keepers in your visualization. Feel free to make the plot as basic as you want, but be sure that the plot makes sense and is clean.

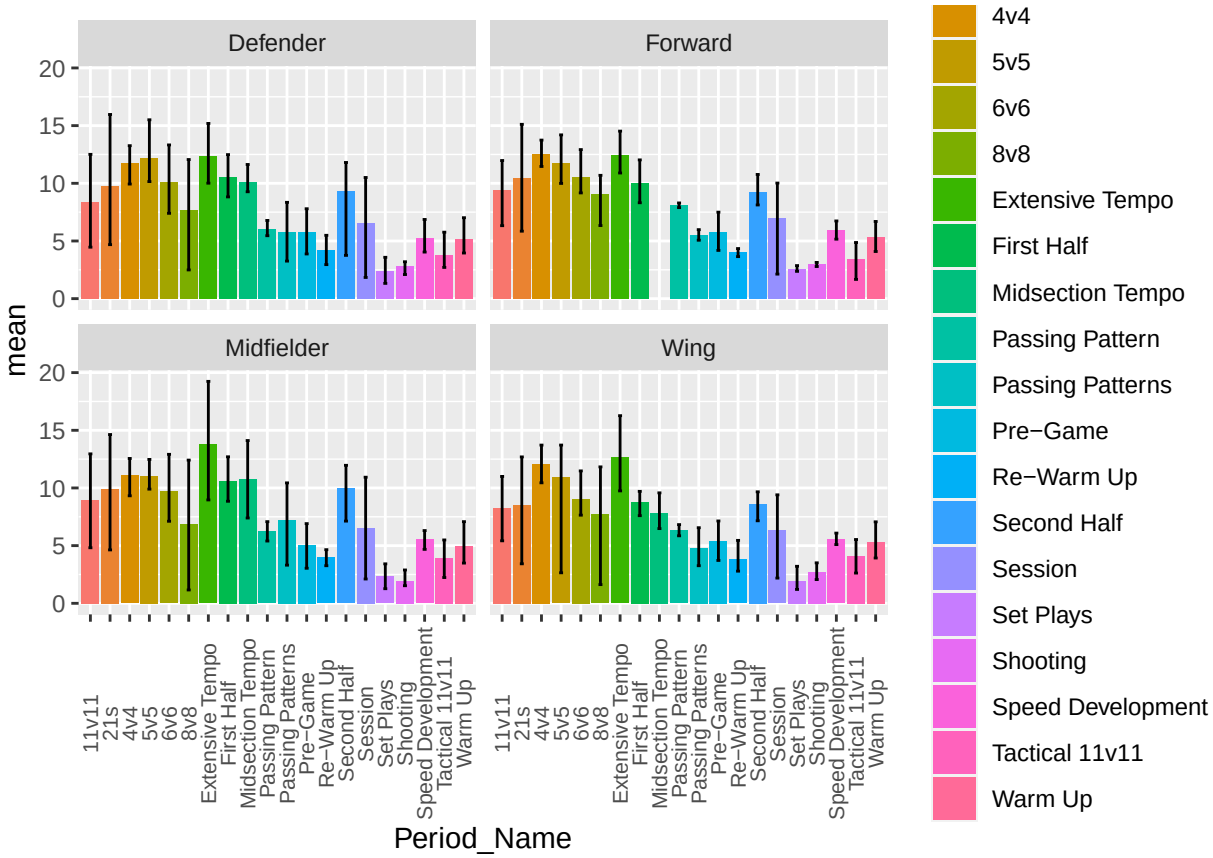
Answer

```
drills = dat %>%
  select(Period_Name, Position_Name, Player_Load_Per_Minute) %>%
  group_by(Period_Name) %>%
```

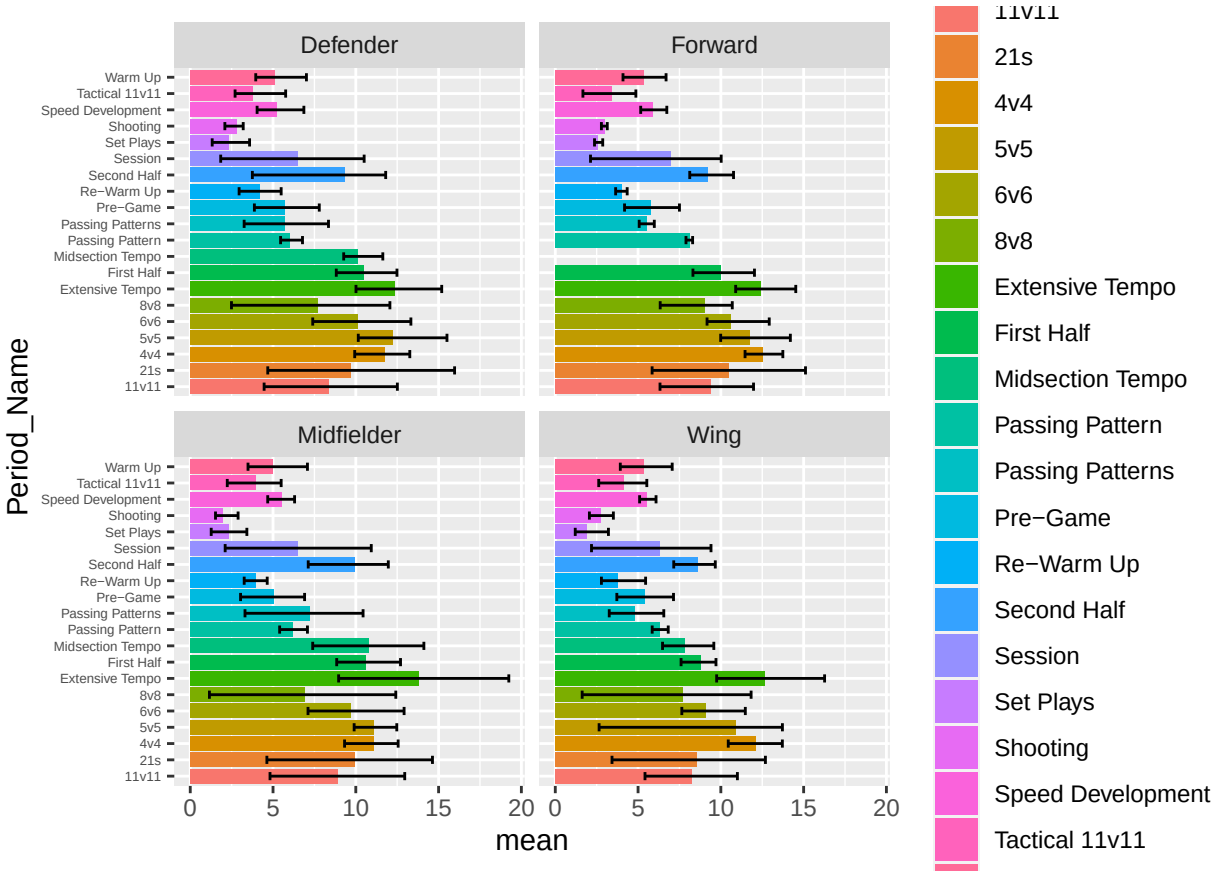
```
  filter(n() >= 20) %>%
  ungroup() %>%
  group_by(Position_Name, Period_Name) %>%
  summarise_at(vars(Player_Load_Per_Minute), list(mean = mean,
                                                  lower = ~ quantile(.,probs = 0.025),
                                                  upper = ~ quantile(.,probs = 0.975)))

drills = drills %>%
  filter(Position_Name != "Goal Keeper")

ggplot(drills, aes(x = Period_Name, y = mean, fill = Period_Name)) +
  geom_bar(stat="identity", position=position_dodge()) +
  geom_errorbar(aes(ymin = lower, ymax = upper), width=.2, position=position_dodge()) +
  facet_wrap( ~ Position_Name) +
  theme(axis.text.x =element_text(angle = 90, vjust = 0.5, size = 8))
```




```
ggplot(drills, aes(y = Period_Name, x = mean, fill = Period_Name)) +
  geom_bar(stat="identity", position=position_dodge()) +
  geom_errorbar(aes(xmin = lower, xmax = upper), width=.5, position=position_dodge()) +
  facet_wrap( ~ Position_Name) +
  theme(axis.text.y = element_text(size=5))
```



Question 3

There is a good chance that the figures you created for Questions 1 and 2 are different than mine. I want you to describe your figures, being sure to break them down so that a coach can understand what they are seeing. However, you have do this in 5 or less sentences per figure.

Answer - Figure 1

Figure 1 visualizes the relative effort a player put into each practice drill on the August 6th practice with respect to Max Velocity, Total Distance, and Total Player Load. The x axis has letters corresponding to each player. The y axis measures their relative effort according to our three metrics. For our metrics, red is max velocity, green is total distance, and blue is total player load. #### Answer - Figure 2 Figure 2 visualizes the average player load per minute for the different drills frequently ran during practice. Each graph is broken down by position, excluding goal keepers. The y axis shows which training drill is being measured. The x shows the player load per minute. The black bars lying within the player load per minute measurements show the confidence interval, with wider bars meaning we are less sure the value displayed matches the true value.