ggplot2 mini

2023-11-20

Part 1 & 2 – Set-up & Starter code, respectively

Load library

```
library(ggplot2)
```

Read in the weather highs by season data from the Grouped Bar Plot module

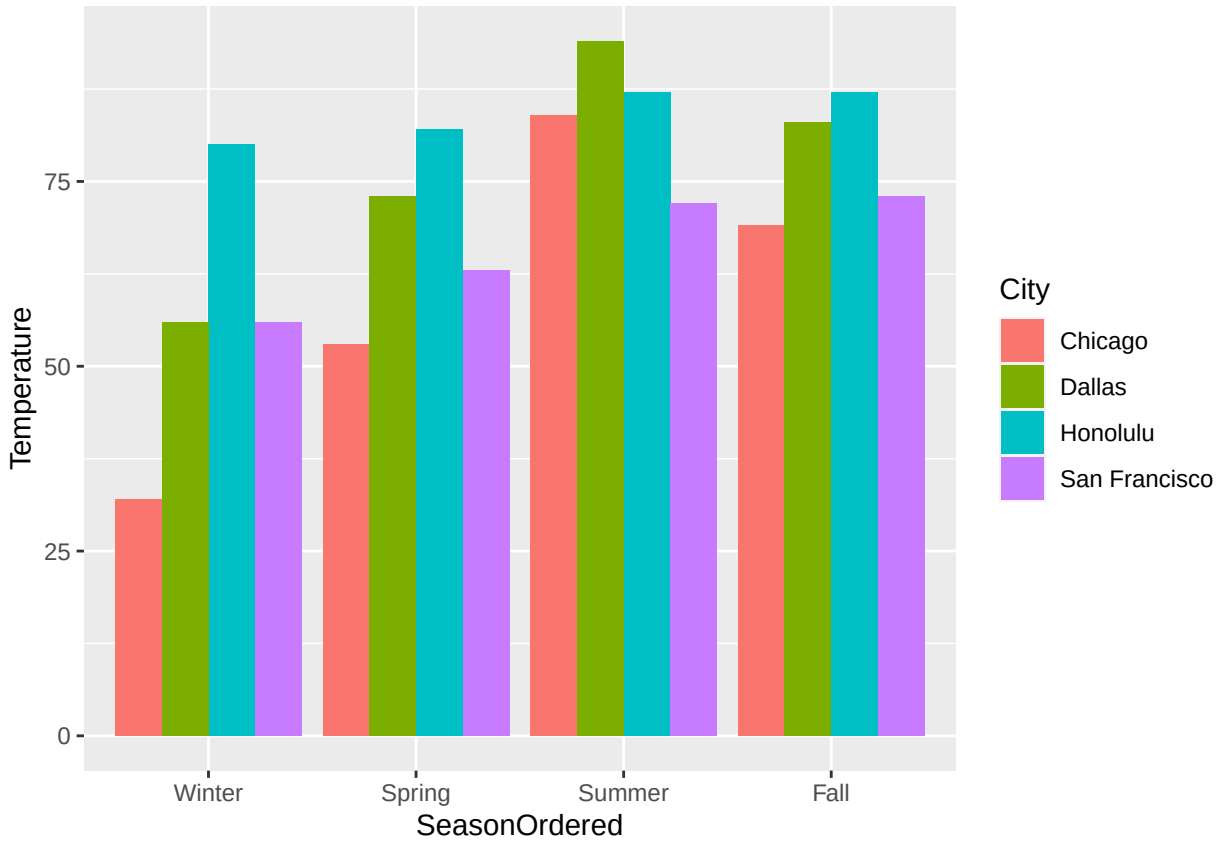
```
df <- read.csv("weather_high_temps.csv")
```

Creating another variable for making the seasons be properly ordered

```
df$SeasonOrdered = factor(df$Season,levels=df$Season[c(1,2,3,4)])
```

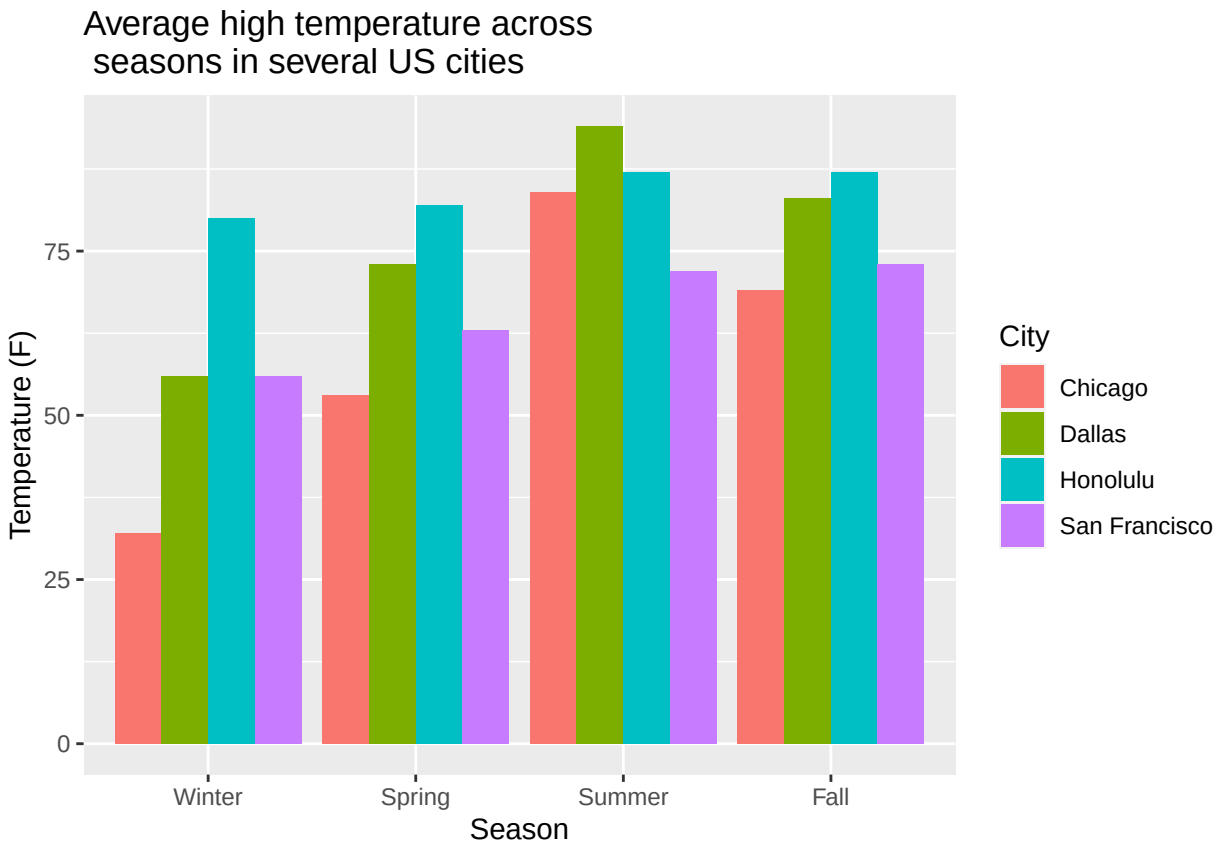
bar (column) plot without labels

```
ggplot(df,aes(x=SeasonOrdered,y=Temperature,fill=City)) +
  geom_col(position="dodge")
```



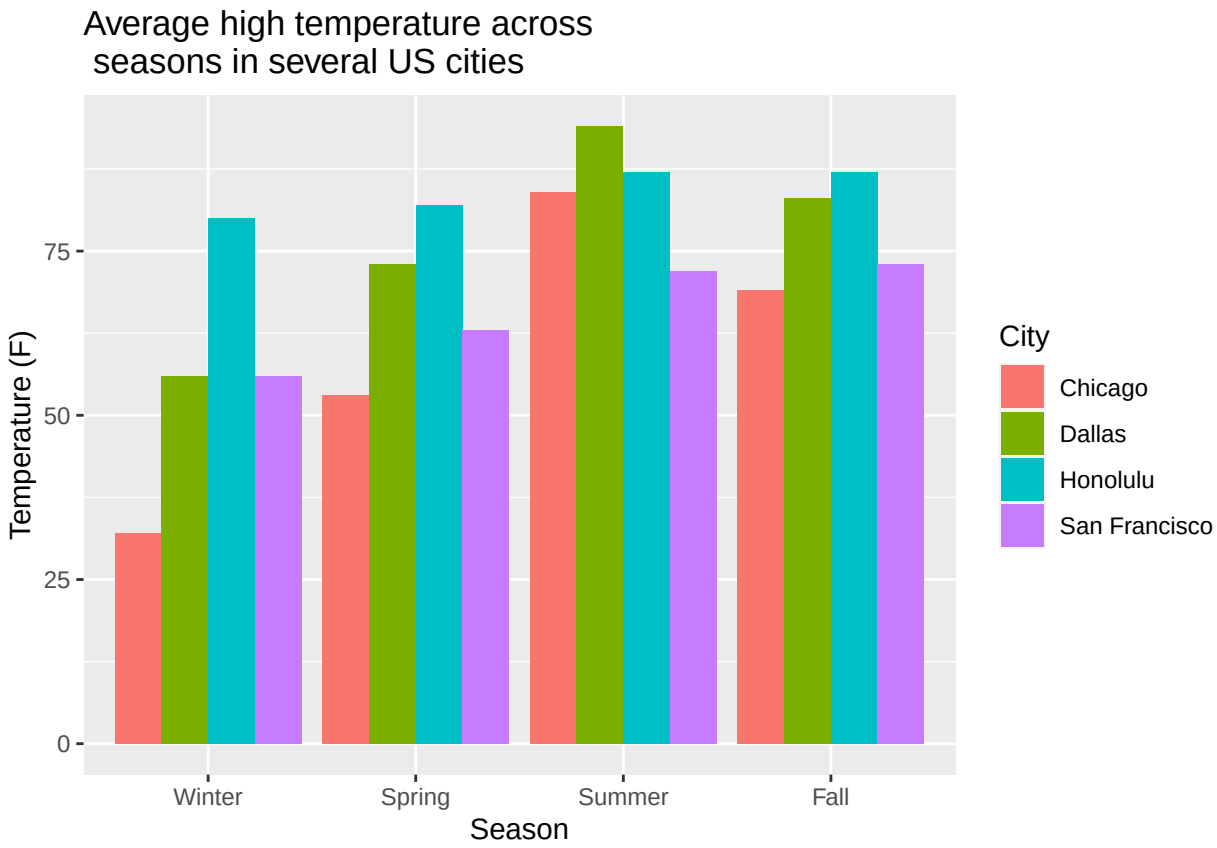
bar plot with labels

```
ggplot(df,aes(x=SeasonOrdered,y=Temperature,fill=City)) +
  geom_col(position="dodge") +
  ggtitle("Average high temperature across \n seasons in several US cities") +
  xlab("Season") +
  ylab("Temperature (F)")
```



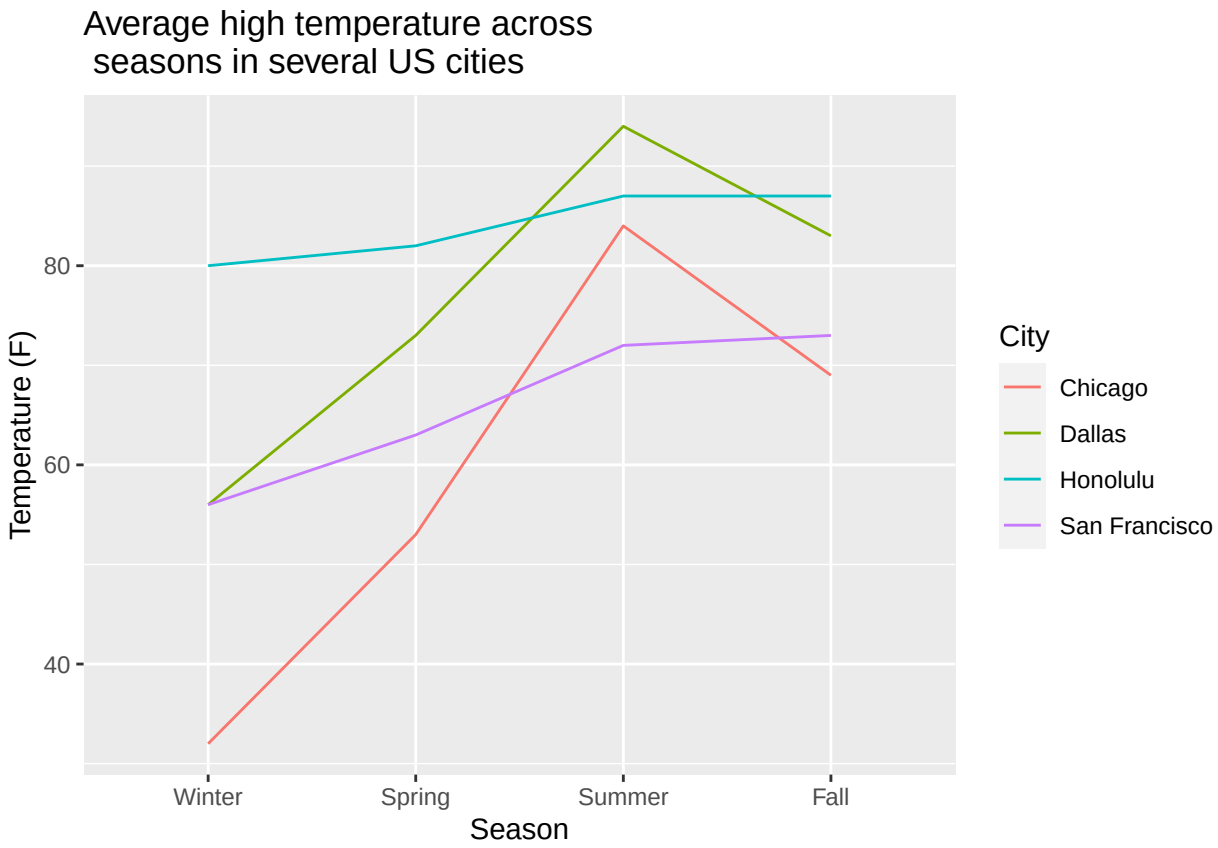
alternate way of creating labels

```
ggplot(df,aes(x=SeasonOrdered,y=Temperature,fill=City)) +
  geom_col(position="dodge") +
  labs(title="Average high temperature across \n seasons in several US cities",x="Season",y="Temperatur
```



line plot

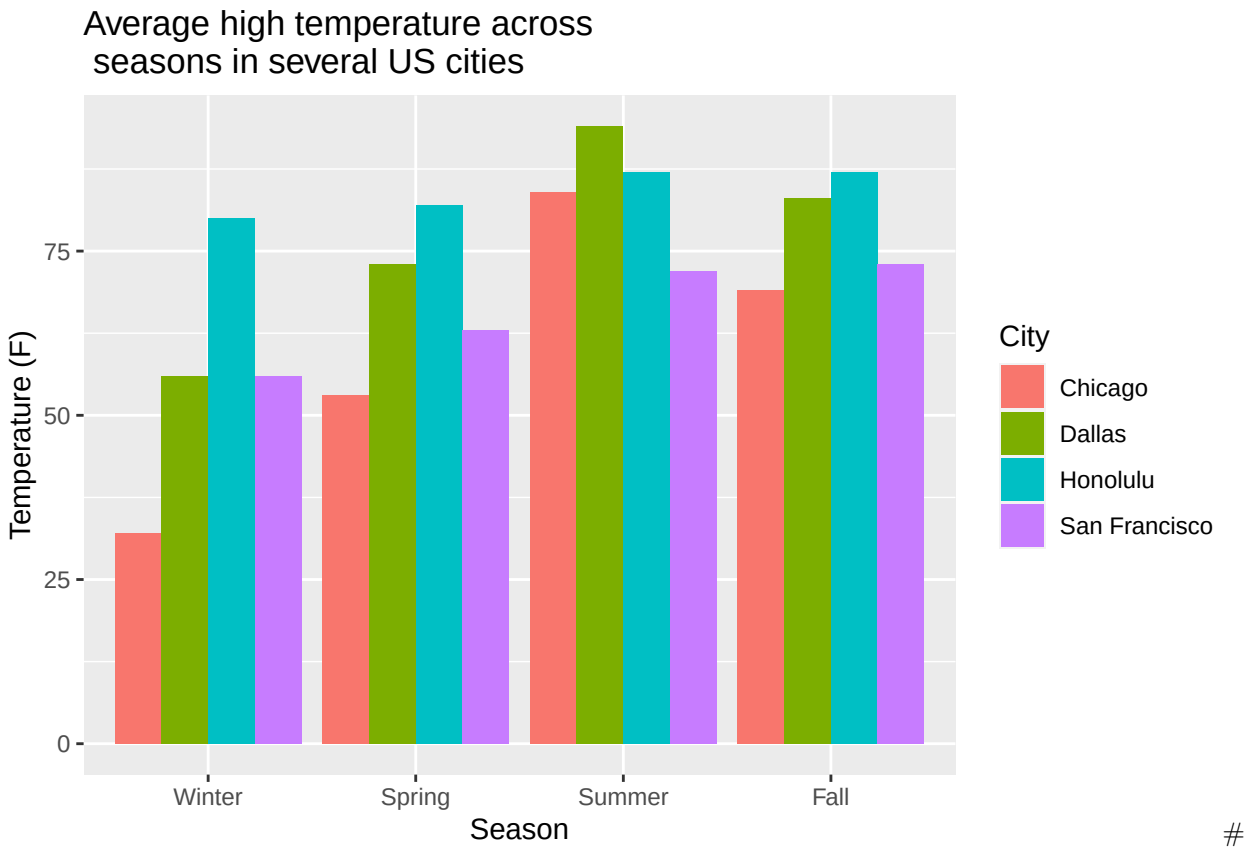
```
ggplot(df,aes(x=SeasonOrdered,y=Temperature,color=City,group=City)) +
  geom_line() +
  ggtitle("Average high temperature across \n seasons in several US cities") +
  xlab("Season") +
  ylab("Temperature (F)")
```



Part 3 – Modify commands in bar plot

Find the part of the code with the bar plot (toward the beginning). It looks like this:

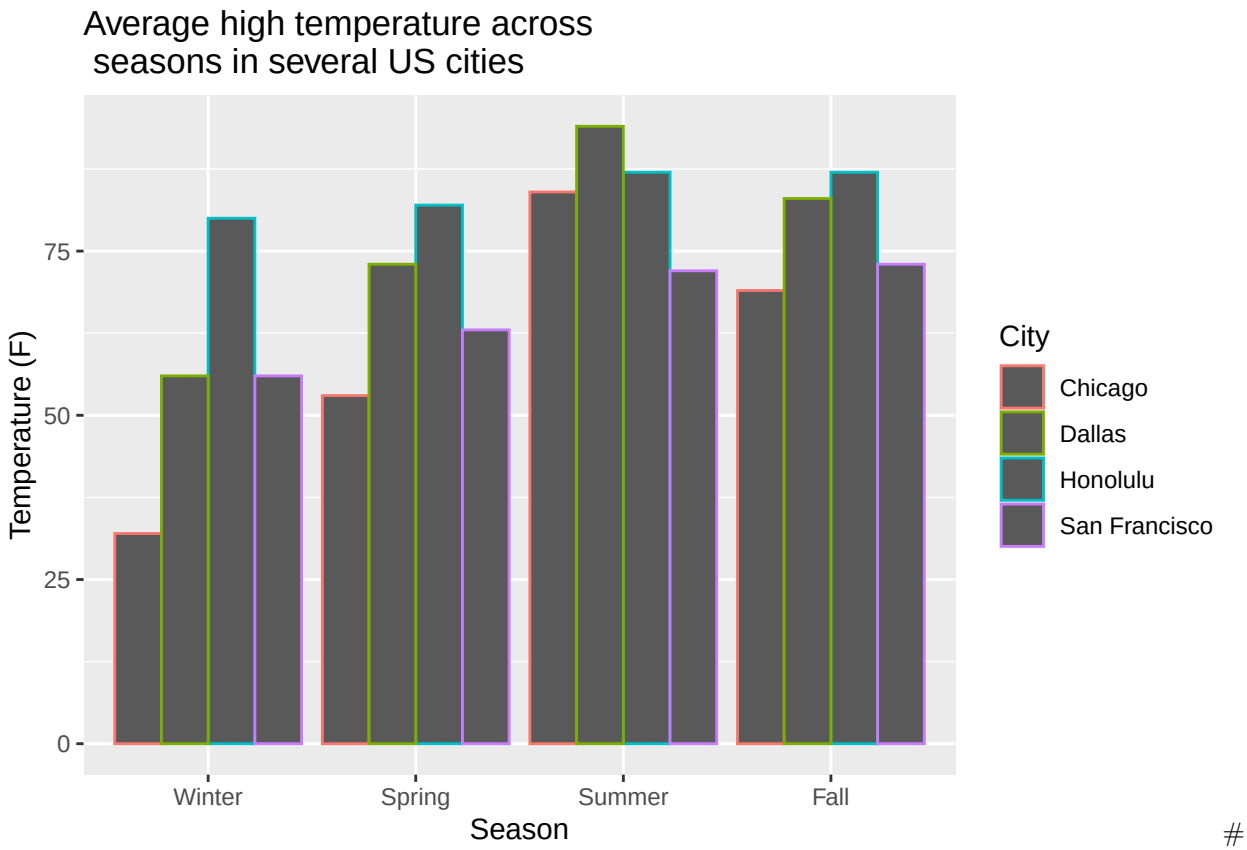
```
ggplot(df,aes(x=SeasonOrdered,y=Temperature,fill=City)) +
   geom_col(position="dodge") +
   ggtitle("Average high temperature across \n seasons in several US cities") +
   xlab("Season") +
   ylab("Temperature (F)")
```



#

1. (*Response) Run the code, then change "fill" to "color" on a bar plot and run it again to see the difference. What does fill do? What does color do?

```
ggplot(df,aes(x=SeasonOrdered,y=Temperature,color=City)) +
   geom_col(position="dodge") +
   ggtitle("Average high temperature across \n seasons in several US cities") +
   xlab("Season") +
   ylab("Temperature (F)")
```

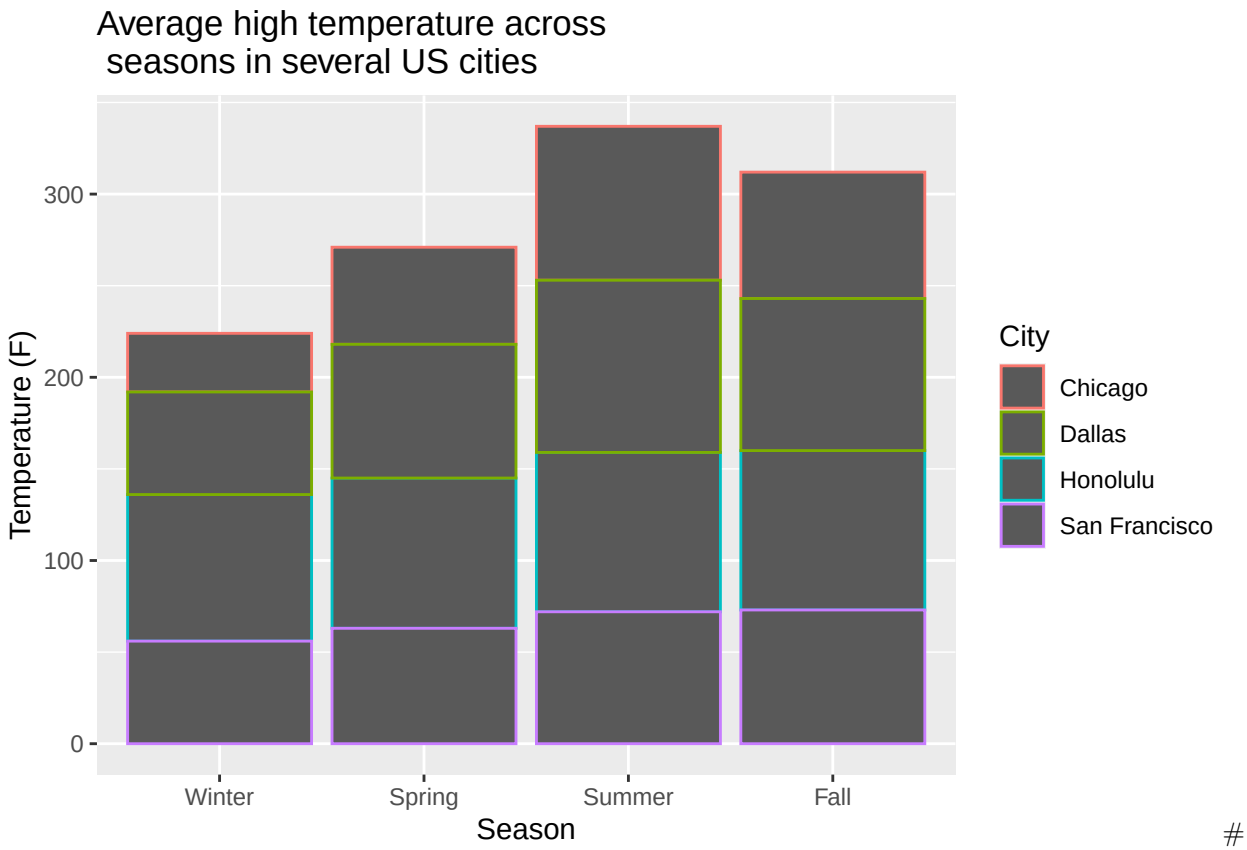


#

Response 1: When you replace "fill" with "color", the colors of the bar charts turn to uniformed colors (i.e., gray), while keeping their individual colored outlined around each bar in accordance to their city. Conversely, when you replace "color" with "fill", the entire bars turn the color of their city, without any grey colors.

2. (*Response) What if you delete the "position = "dodge"" - what's the difference? What does "position = "dodge"" accomplish?

```
ggplot(df,aes(x=SeasonOrdered,y=Temperature,color=City)) +
   geom_col() +
   ggtitle("Average high temperature across \n seasons in several US cities") +
   xlab("Season") +
   ylab("Temperature (F)")
```

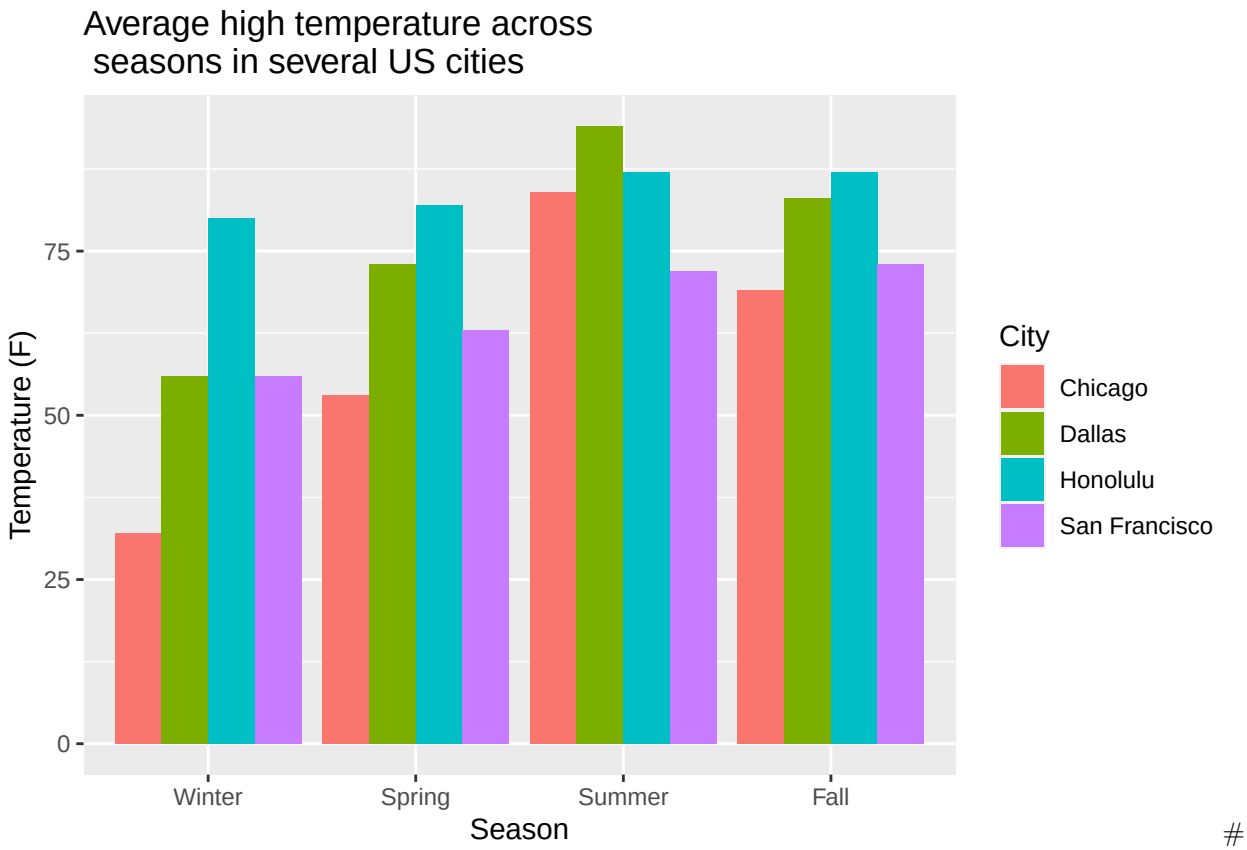


#

Response 2: When you delete "position="dodge" " in the code, the x-axis is visualized as a stacked bar plot instead of graphed bars next to each other. As a result, the y-axis's temperature values are higher to reflect the total stacked values for each season.

3. (*Response) Go back to the original code and delete the + sign on the line before the "ggtitle" command. Put your cursor in the first two lines of the command and run it. What does the plot look like (describe the problem)? In what case might you look for a bug in your code where a + sign was forgotten?

```
ggplot(df,aes(x=SeasonOrdered,y=Temperature,fill=City)) +
   geom_col(position="dodge") +     #remove +
   ggtitle("Average high temperature across \n seasons in several US cities") +
   xlab("Season") +
   ylab("Temperature (F)")
```



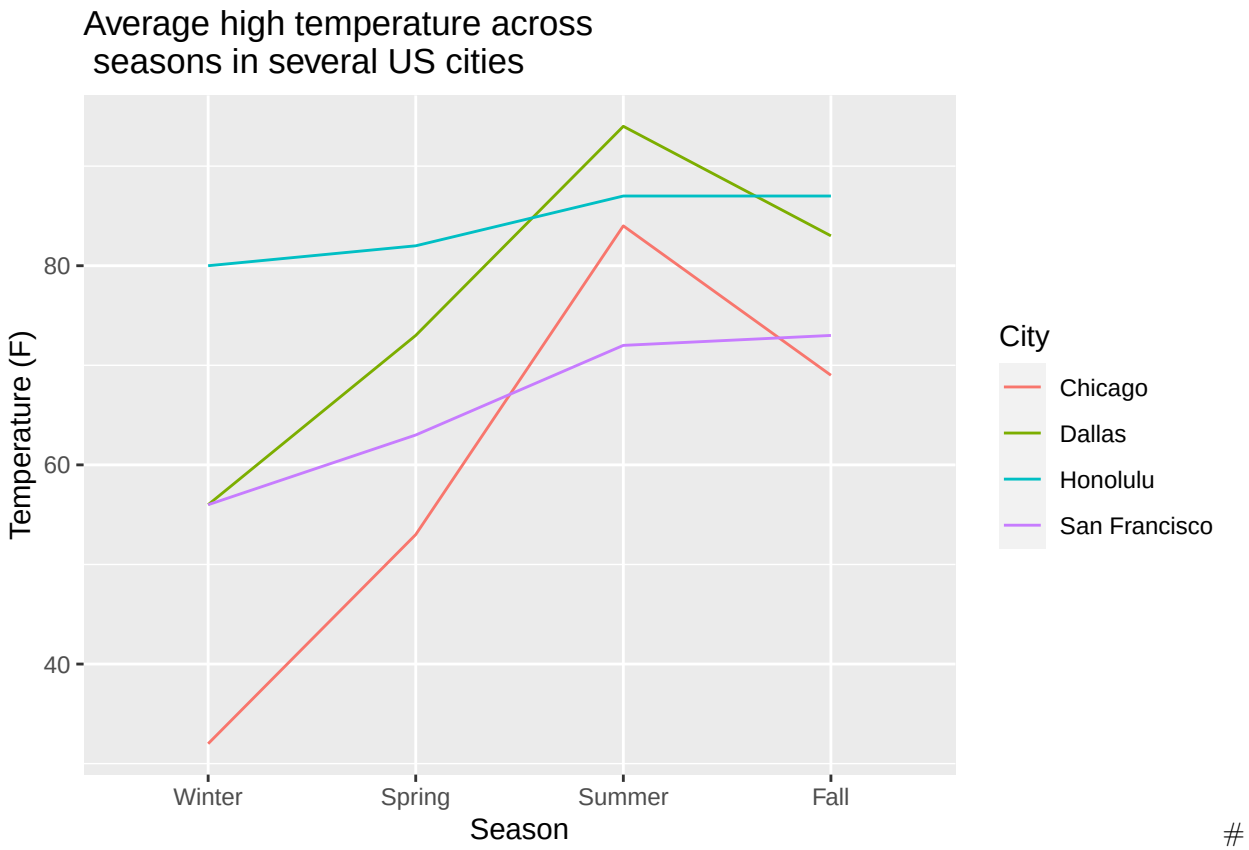
#

Response 3: By removing the "+", it takes away the plot's title and redescribes the x-axis as "SeasonsOrdered". I think the columns and plot title having errors would be an indicator that there might be a "+" missing, or any plot where there was missing necessities (e.g., variable labels).

Part 4 – Modify commands in line plot

Now find the basic line plot. The code looks like this:

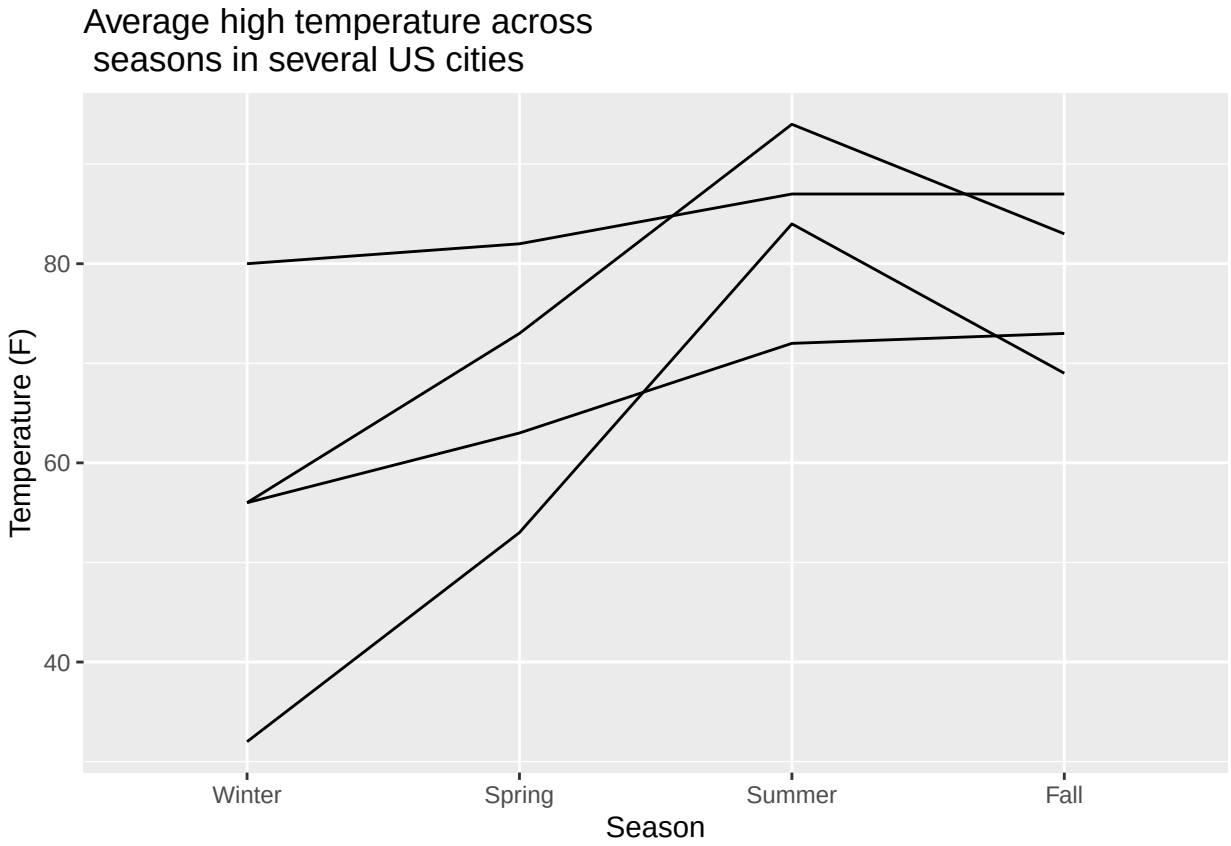
```
ggplot(df,aes(x=SeasonOrdered,y=Temperature,color=City,group=City)) +
   geom_line() +
   ggtitle("Average high temperature across \n seasons in several US cities") +
   xlab("Season") +
   ylab("Temperature (F)")
```



#

1. (*Response) Remove the "color" option on the line plot by deleting this part (including the comma): "color=City,". What happens?

```
ggplot(df,aes(x=SeasonOrdered,y=Temperature,group=City)) +
    geom_line() +
    ggtitle("Average high temperature across \n seasons in several US cities") +
    xlab("Season") +
    ylab("Temperature (F)")
```



Response 1: The colors of the lines are removed, so you cannot tell which city each line corresponds to. There is also no legend.

2. (*Response) Now, copy and paste the code for the original basic line plot provided above, and replace the word "color" with "shape". Also add in a point geom. The first line should now look like this:

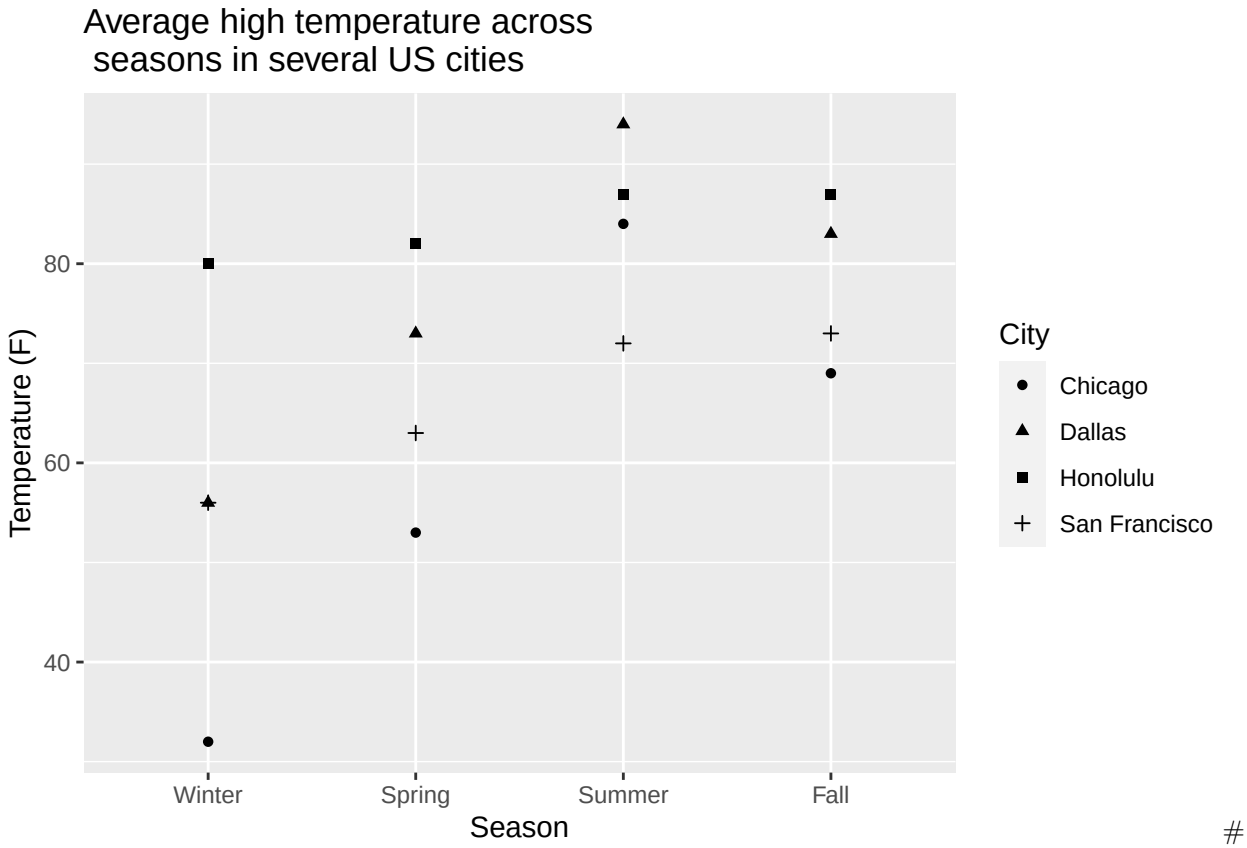
ggplot(df,aes(x=SeasonOrdered,y=Temperature,shape=City,group=City)) +

At the end of the first line, include the code:

geom_point() +

What is the result?

```
ggplot(df,aes(x=SeasonOrdered,y=Temperature,shape=City,group=City)) +
  geom_point() +
  ggtitle("Average high temperature across \n seasons in several US cities") +
  xlab("Season") +
  ylab("Temperature (F)")
```



#

Response 2: The cities are each now visualized by shapes, with each city reflected once along the x-axis (except San Francisco is not seen on x-axis "Winter").

3. (*Response) For you, which is easier to interpret, this or the color plot?

Response 3: The color plot is easiest to interpret simply because the colors are each very different per city, and can be seen in the legend. It is also interesting to see a line to represent a trend in change per city.

4. (*Code) Fortunately, we don't have to limit ourselves to one plot type or another; we can use multiple layers. Now let's create a line plot with color, shape, and linetype all coding the City, which will make the lines very easy to distinguish. Include both the "color=City" and "shape=City" components. Be sure to have exactly one comma between each of the items in the parentheses. Include both the geom_line() and geom_point() geometries. Run the code to make sure you can make a working plot with both line colors and point shapes before moving on.

Next, go to this official ggplot2 documentation page (https://ggplot2.tidyverse.org/reference/aes_group_order.html) and search for the word "linetype". Follow the pattern demonstrated to make your code create a plot with a different linetype for each city in addition to the existing color and shape markers.

```
# Response 4: linetype
ggplot(df,aes(x=SeasonOrdered,y=Temperature,color=City,shape=City)) +
  geom_line(aes(linetype = SeasonOrdered)) +
  geom_point() +
  ggtitle("Average high temperature across \n seasons in several US cities") +
  xlab("Season") +
  ylab("Temperature (F)")
```

```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```