

## Problem Situation – Part II

### Data Visualization

16 tasks to be completed.

1. Change the name to the crunchy\_wages variables (in case you didn't before) and replace Global by USA in both data sets (crunchy\_wages and crunchy\_sales).

```
crunchy_wages <- crunchy_wages %>%
  rename(Number_of_workers= `Number of workers` ) %>%
  rename(Wage_USD= `Wage (USD)` ) %>%
  mutate(Country=replace(Country,Country=="USA", "Global"))
```

2. Get the names for the levels of the Month variable in crunchy\_sales. You need to change everything to english language. Verify at the end (with function levels) that this was solved.

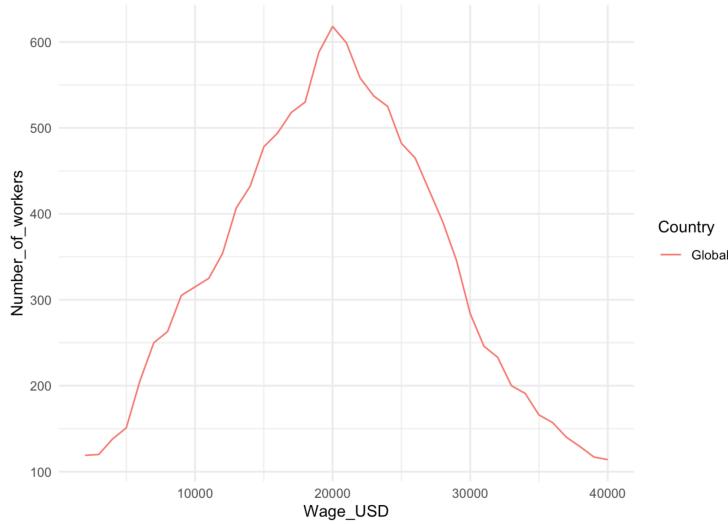
```
crunchy_sales <- crunchy_sales %>%
  mutate(Month=replace(Month,Month=="Enero", "January")) %>%
  mutate(Month=replace(Month,Month=="Febrero", "February")) %>%
  mutate(Month=replace(Month,Month=="Marzo", "March")) %>%
  mutate(Month=replace(Month,Month=="Abril", "April")) %>%
  mutate(Month=replace(Month,Month=="Mayo", "May")) %>%
  mutate(Month=replace(Month,Month=="Junio", "June")) %>%
  mutate(Month=replace(Month,Month=="Julio", "July")) %>%
  mutate(Month=replace(Month,Month=="Agosto", "August")) %>%
  mutate(Month=replace(Month,Month=="Sepbre", "September")) %>%
  mutate(Month=replace(Month,Month=="Octubre", "October")) %>%
  mutate(Month=replace(Month,Month=="Noviembre", "November")) %>%
  mutate(Month=replace(Month,Month=="Diciembre", "December"))

levels(as.factor(crunchy_sales$Month))
```

3. Replace “Región 1” by “Region 1” and so on in crunchy\_sales.

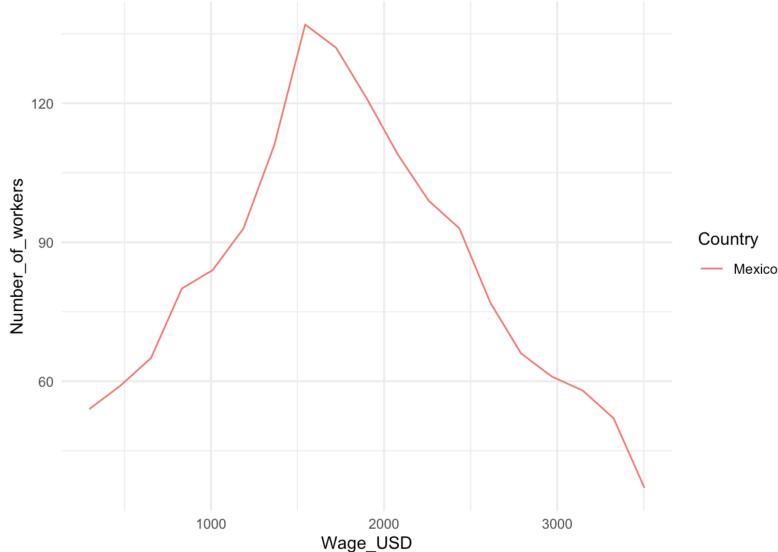
```
crunchy_sales <- crunchy_sales %>%
  mutate(Region=replace(Region,Region=="Región 1", "Region 1")) %>%
  mutate(Region=replace(Region,Region=="Región 2", "Region 2")) %>%
  mutate(Region=replace(Region,Region=="Región 3", "Region 3")) %>%
  mutate(Country=replace(Country,Country=="USA", "Global"))
```

4. Create the following graph for Global wages. Change the name of the axis labels to Number of workers and Wage (USD).



```
crunchy_wages %>%
  filter(Country=="Global") %>%
  ggplot(aes(x=Wage_USD,y=Number_of_workers))+geom_line(aes(col=Country))+theme_minimal()
```

5. Create the previous graph now for Mexico wages. Change the name of the axis labels to Number of workers and Wage (USD).



**Question to be answered in your final report (Part III of the problem situation) will be marked on blue.**  
 What distribution seems to have each one of the graphs? Do you notice skewness? What recommendation would you provide related to the wages of Crunchy Co.

6. Build the following table, it shows the total sales for every product by country. What conclusion would you get from it? Do the products perform the same in Mexico compared to other countries?

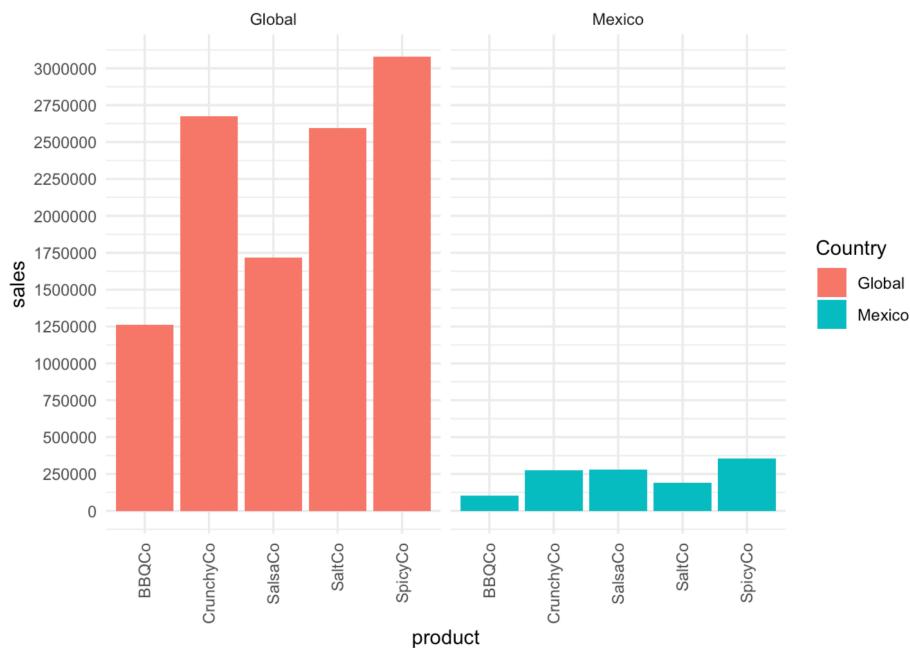
Country	CrunchyCo	SpicyCo	SaltCo	BBQCo	SalsaCo
Global	2,675,941.0	3,076,885	2,596,543.0	1,263,733.0	1,717,814
Mexico	273,772.8	355,172	189,607.8	101,331.5	277,986

```

sales_by_country <- crunchy_sales %>%
  select(-Row) %>%
  group_by(Country) %>%
  summarise_if(is.numeric,(sum))
kable(sales_by_country,format.args = list(big.mark = ","))

```

7. Build the following graph that represents the previous table. Use minimal theme. In your final report, it must have an interpretation.



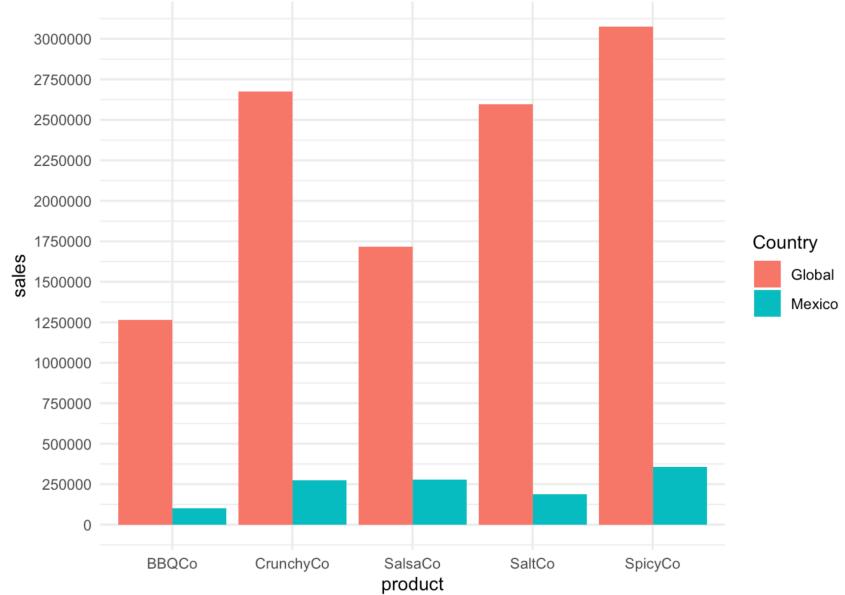
```

total_sales <- crunchy_sales %>%
  select(-c(Row,Month,Region))

total_sales_long <- total_sales %>%
  pivot_longer(CrunchyCo:SalsaCo, names_to = "product", values_to = "sales")
options(scipen=999)
ggplot(total_sales_long,aes(x=product,y=sales,fill=Country))+geom_col()+facet_wrap(~Country)+theme(axis.text.x =
  element_text(angle = 90, vjust = 0.5, hjust=1))+scale_y_continuous(breaks=seq(0,3100000,250000))+theme_minimal()+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))

```

8. Build the following dodge bar plot. In your final report, it must have an interpretation. Which one seems to be better to represent the dataset?



```
ggplot(total_sales_long,aes(x=product,y=sales,fill=Country))+stat_summary(geom="col",fun="sum",position="dodge")+
  scale_y_continuous(breaks=seq(0,3100000,250000))+theme_minimal()
```

9. Now, let's focus on the Mexican market. Build the following table. What can you say about the performance of each region/product?

Region	CrunchyCo	SpicyCo	SaltCo	BBQCo	SalsaCo
Region 1	93,406.0	111,825	60,934.88	34,901.76	92,529
Region 2	83,650.0	118,768	60,420.88	34,675.76	97,443
Region 3	96,716.8	124,579	68,252.00	31,754.00	88,014

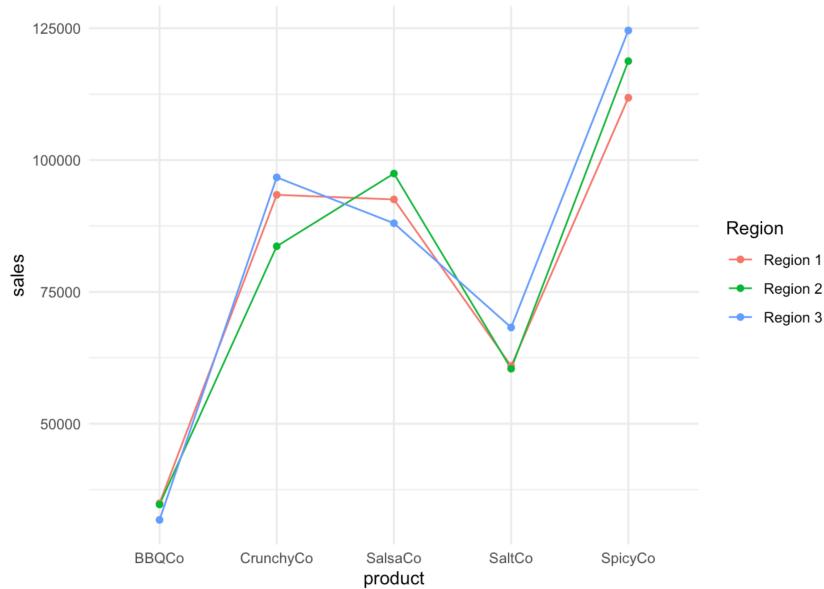
```

regions_mexico <- crunchy_sales %>%
  select(-Row) %>%
  filter(Country=="Mexico") %>%
  group_by(Region) %>%
  summarise_if(is.numeric,(sum))

kable(regions_mexico,format.args = list(big.mark = ","))

```

10. Build a graph to represent the previous table. What recommendations would you provide based on it? Also, explain in detail, in your own words, what is the pivot\_longer function doing.



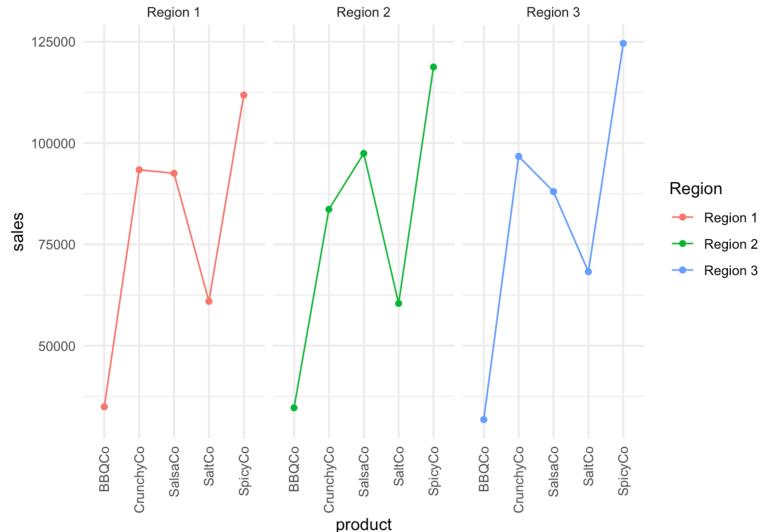
```

region_mexico_long <- regions_mexico %>%
  pivot_longer(CrunchyCo:SalsaCo, names_to = "product", values_to = "sales")

region_mexico_long %>%
  ggplot(aes(x=product,y=sales,col=Region,group=Region))+geom_line()+geom_point()+theme_minimal()

```

11. Now, build separated graphs por each Mexican region using facets.



```
region_mexico_long %>%
  ggplot(aes(x=product,y=sales,col=Region,group=Region))+geom_line()+geom_point()+theme_minimal()+facet_wrap(~Region)+theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

12. Build a table that summarizes the performance for each region in Mexico.

Region	Total_sales
Region 1	393,596.6
Region 2	394,957.6
Region 3	409,315.8

```
sales_by_region_mx <- region_mexico_long %>%
  group_by(Region) %>%
  summarise(Total_sales=sum(sales))
kable(sales_by_region_mx,format.args = list(big.mark = ", " ),align = rep("c",2))
```

13. Build a table that summarizes the total sales for product in Mexico.

product	Total_sales
BBQCo	101,331.5
CrunchyCo	273,772.8
SalsaCo	277,986.0
SaltCo	189,607.8
SpicyCo	355,172.0

```
sales_by_product_mx <- region_mexico_long %>%
  group_by(product) %>%
  summarise(Total_sales=sum(sales))
kable(sales_by_product_mx,format.args = list(big.mark = ", " ),align = rep("c",2))
```

14. Prepare a table that shows the total sales for each month in Mexico. Which seems to be the best and worst month for sales?

Month	Total Sales
January	73,450.29
February	90,156.00
March	108,379.00
April	102,801.00
May	101,238.00
June	99,870.00
July	103,972.00
August	109,566.00
September	96,267.00
October	109,489.00
November	102,586.00
December	100,095.80

```

months_mexico <- crunchy_sales %>%
  select(-Row) %>%
  filter(Country=="Mexico") %>%
  group_by(Month) %>%
  summarise_if(is.numeric,(sum))

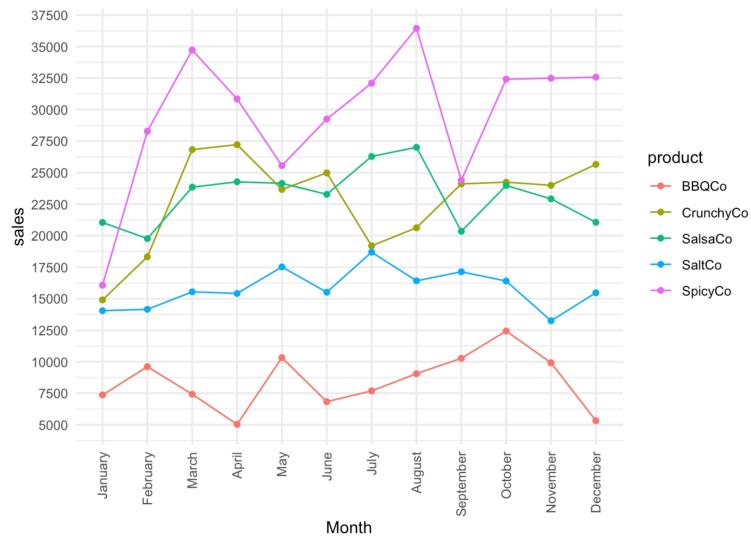
months_mexico_long <- months_mexico %>%
  pivot_longer(CrunchyCo:SalsaCo, names_to = "product", values_to = "sales")

sales_by_month_mx <- months_mexico_long %>%
  group_by(Month) %>%
  mutate(Month = factor(Month, levels = month.name)) %>%
  summarise("Total Sales"=sum(sales))

kable(sales_by_month_mx,format.args = list(big.mark = ","),align = rep("c",2))

```

15. Focus now on the sales for each product by month, build a graph to show them.



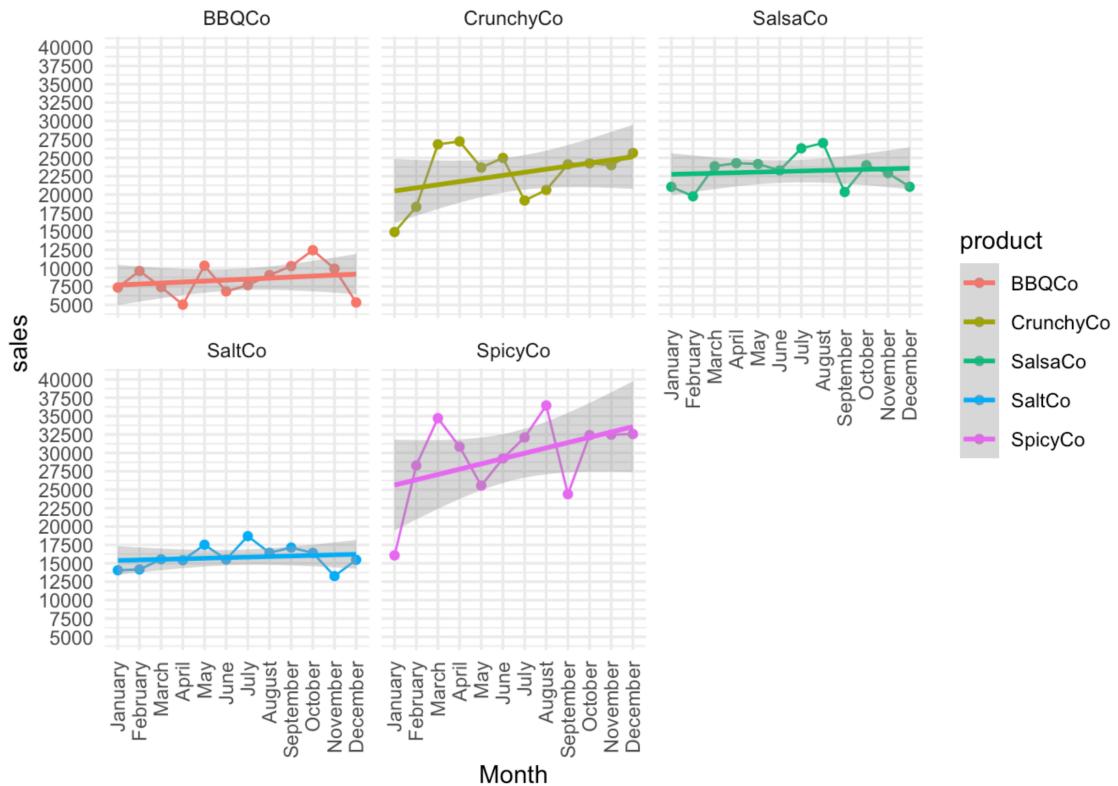
```

months_mexico_long %>%
  ggplot(aes(x=Month,y=sales,col=product,group=product))+geom_line()+geom_point()+theme_minimal()+theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+ scale_x_discrete(limits = month.name)+scale_y_continuous(breaks=seq(0,40000,2500))

```

16. Finally, build a graph that shows individually the sales by month for each product, and add a trend line.

What are your findings? What recommendations would you provide based on it?



```
months_mexico_long %>%
  ggplot(aes(x=Month,y=sales,col=product,group=product))+geom_line()+geom_point()+theme_minimal()+theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+ scale_x_discrete(limits = month.name)+scale_y_continuous(breaks=seq(0,40000,2500))+facet_wrap(~product)+geom_smooth(method = lm)
```