Data Wrangling – Part 4

`summarize()` Function

Command Illustration

```
new_dataframe_name <- dataframe_name %>%
      summarise(new_column_name = function_name(column_name2))
```

group_by() & summarize() Functions

Command Illustration

```
new_dataframe_name <- dataframe_name %>%
      group_by(column_name1) %>%
      summarise(new_column_name = function_name(column_name2))
```

Illustration_Data

Name	Age	total_Income	var_1	var_2	var_3	zipcode	honesty	cat_total
Val	18	18000	apple	carrots	elephant	60001	agree	0
Derek	25	25000	grapes	carrots	tiger	60073	disagree	1
Whitney	30	30000	bananas	carrots	lion	60109	disagree	2
Sasha	40	40000	peaches	carrots	rabbit	60111	disagree	1
Daniella	45	45000	bananas	carrots	shark	60155	agree	1

Example 1: Compute mean and median of total_Income

```
example_1 <- Illustration_Data %>%
      summarise(mean_income = mean(total_Income),
                median_income = median(total_Income))
```

Name	Age	total_Income	var_1	var_2	var_3	zipcode	honesty	cat_total
Val	18	18000	apple	carrots	elephant	60001	agree	0
Derek	25	25000	grapes	carrots	tiger	60073	disagree	1
Whitney	30	30000	bananas	carrots	lion	60109	disagree	2
Sasha	40	40000	peaches	carrots	rabbit	60111	disagree	1
Daniella	45	45000	bananas	carrots	shark	60155	agree	1

Example 1: Compute mean and median of total_Income

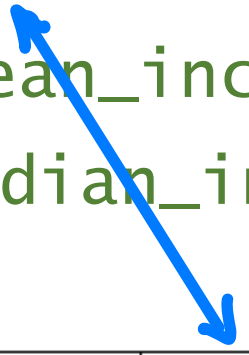
```
example_1 <- Illustration_Data %>%
      summarise(mean_income = mean(total_Income),
                median_income = median(total_Income))
```

Output: The output is a dataframe that looks like this

mean_income	median_income
31,600	30,000

Example 2: Compute mean and median of total_Income, grouped by favorite fruit (var_1)

```
example_2 <- Illustration_Data %>%
      group_by(var_1) %>%
      summarise(mean_income = mean(total_Income),
                median_income = median(total_Income))
```



Name	Age	total_Income	var_1	var_2	var_3	zipcode	honesty	cat_total
Val	18	18000	apple	carrots	elephant	60001	agree	0
Whitney	30	30000	bananas	carrots	lion	60109	disagree	2
Daniella	45	45000	bananas	carrots	shark	60155	agree	1
Derek	25	25000	grapes	carrots	tiger	60073	disagree	1
Sasha	40	40000	peaches	carrots	rabbit	60111	disagree	1

Example 2: Compute mean and median of total_Income, grouped by favorite fruit (var_1)

```
example_2 <- Illustration_Data %>%
      group_by(var_1) %>%
      summarise(mean_income = mean(total_Income),
                median_income = median(total_Income))
```

Output: The output is a dataframe that looks like this

var_1	mean_income	median_income
apple	18,000	18,000
bananas	37,500	37,500
grapes	25,000	25,000
peaches	40,000	40,000

Example 3: Compute sum and mean of total_Income, grouped by honesty (var_1) and cat_total

```
example_3 <- Illustration_Data %>%
     group_by(honesty, cat_total) %>%
     summarise(sum_income = sum(total_Income),
                    mean_income = mean(total_Income))
```

Name	Age	total_Income	var_1	var_2	var_3	zipcode	honesty	cat_total
Val	18	18000	apple	carrots	elephant	60001	agree	0
Daniella	45	45000	bananas	carrots	shark	60155	agree	1
Derek	25	25000	grapes	carrots	tiger	60073	disagree	1
Sasha	40	40000	peaches	carrots	rabbit	60111	disagree	1
Whitney	30	30000	bananas	carrots	lion	60109	disagree	2

Example 3: Compute sum and mean of total_Income, grouped by honesty (var_1) and cat_total

```
example_3 <- Illustration_Data %>%
      group_by(honesty, cat_total) %>%
      summarise(sum_income = sum(total_Income),
                mean_income = mean(total_Income))
```

Output: The output is a dataframe that looks like this

honesty	cat_total	Sum_income	mean_income
agree	0	18,000	18,000
agree	1	45,000	45,000
disagree	1	65,000	32,500
disagree	2	30,000	30,000

Example 4: Compute the number of pro dancers per honesty

```
example_4 <- Illustration_Data %>%
      group_by(honesty) %>%
      summarise(Num_Pro = n())
```

counts the number of rows

Name	Age	total_Income	var_1	var_2	var_3	zipcode	honesty	cat_total
Val	18	18000	apple	carrots	elephant	60001	agree	0
Daniella	45	45000	bananas	carrots	shark	60155	agree	1
Derek	25	25000	grapes	carrots	tiger	60073	disagree	1
Sasha	40	40000	peaches	carrots	rabbit	60111	disagree	1
Whitney	30	30000	bananas	carrots	lion	60109	disagree	2

Example 4: Compute the number of pro dancers per honesty

```
example_4 <- Illustration_Data %>%
      group_by(honesty) %>%
      summarise(Num_Pro = n())
```

Output: The output is a dataframe that looks like this

honesty	Num_pro
agree	2
disagree	3