Bivariate Graphing Assignment

Pushpak Sunil Rane

Describing Relationship

C ~ C Association

country ~ disagg.value.x

"country" variable represents regions or places for analyzing diet and impact of it by environmental factors.

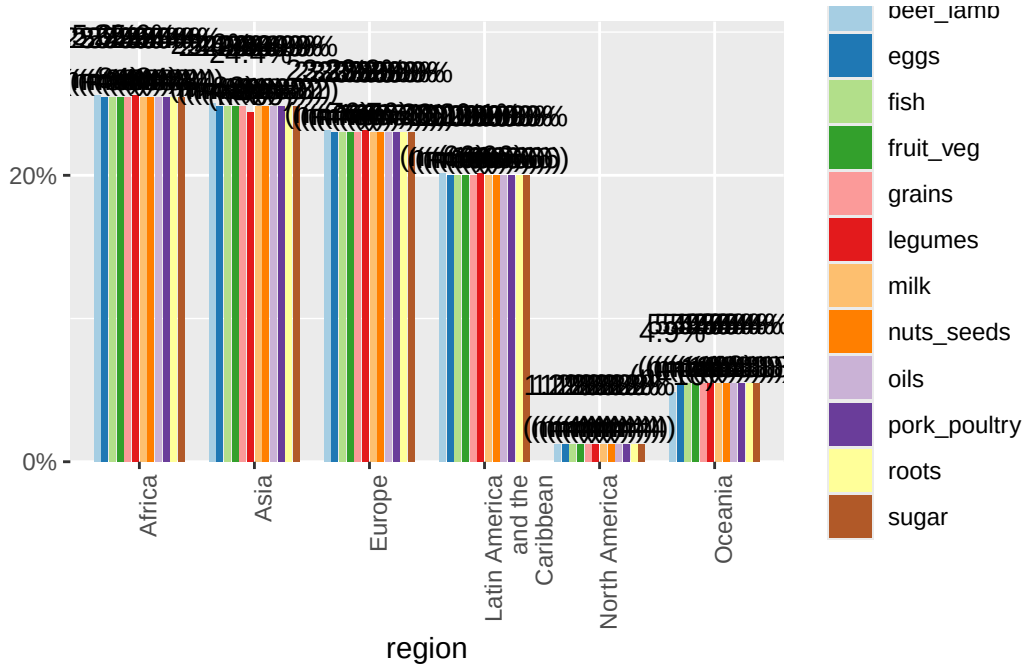
"*disagg.value.x*" variable represents diet items(fruits,vegetables,beef,etc) in particular country.

```
mydata %>% select(region, disagg.value.x) %>%
  tbl_summary(by = "disagg.value.x",
              statistic = list(all_continuous() ~ "{mean} ({sd})"))
```

Characteristic	**beef_lamb** N = 328[1]	**eggs** N = 330[1]	**fish** N = 330[1]	**fru**
region				
Africa	84 (26%)	84 (25%)	84 (25%)	
Asia	82 (25%)	82 (25%)	82 (25%)	
Europe	76 (23%)	76 (23%)	76 (23%)	
Latin America and the Caribbean	66 (20%)	66 (20%)	66 (20%)	
North America	4 (1.2%)	4 (1.2%)	4 (1.2%)	
Oceania	16 (4.9%)	18 (5.5%)	18 (5.5%)	

[1]n (%)

```
plot_xtab(mydata$region, grp = mydata$disagg.value.x, show.total = FALSE) +
theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



The bar chart you provided illustrates the relationship between the variable "region" and the categories within "disagg.value.x". Each bar represents a country, and the different colored segments within each bar represent the proportions of various categories in "disagg.value.x" (e.g., beef_lamb, eggs, fruit_veg, etc.). Approximately 25% - 26% of Africa has higher impact on diet of each item in "disagg.value.x". Whereas, 1.2% of North America has lowest impact.

Q ~ C Association

planetary_impacts_Cropland ~ disagg.value.x

"planetary_impacts_Cropland" variable represents environmental impact of using freshwater (use in %) on diet and nutrition in certain regions.

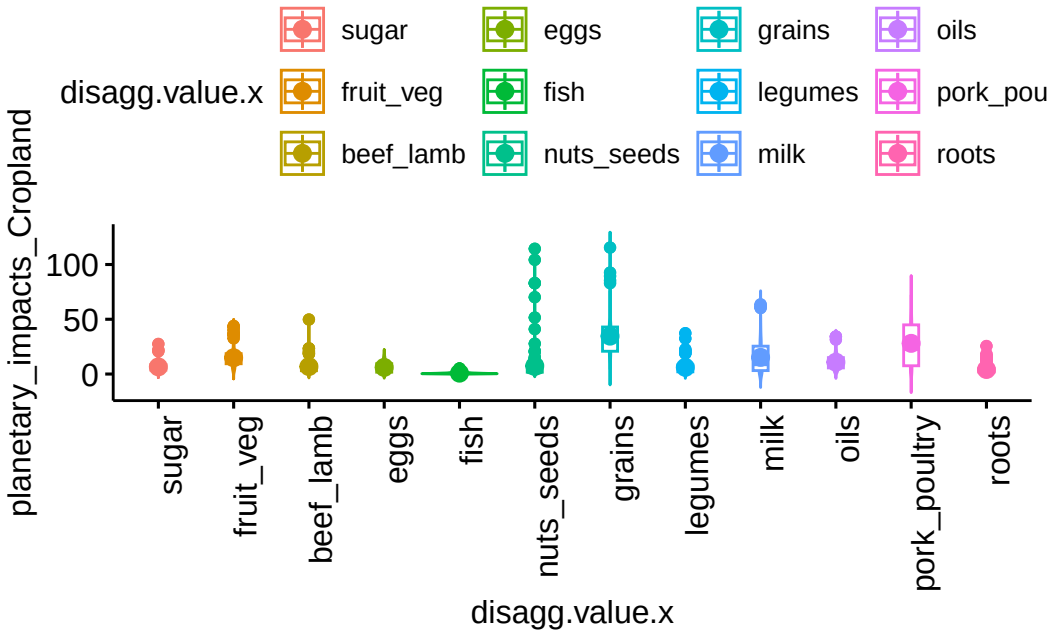
"*disagg.value.x*" variable represents diet items(fruits,vegetables,beef,etc) grams/day in particular country.

```
mydata %>% select(planetary_impacts_Cropland, disagg.value.x) %>%
  tbl_summary(by = "disagg.value.x",
  statistic = list(all_continuous() ~ "{mean} ({sd})"))
```

Characteristic	beef_lamb N = 328[1]	eggs N = 330[1]	fish N = 330[1]	fruit_veg
planetary_impacts_Cropland	7 (6)	6 (5)	1 (1)	15 (

[1]Mean (SD)

```
ggviolin(mydata,
  x="disagg.value.x", y = "planetary_impacts_Cropland",
  color = "disagg.value.x", add = c("mean", "boxplot")) + theme(axis.text.x = element_text(a
```



The distribution of planetary_impacts_Cropland are fairly normal for each disagg.value.x, with some higher average values of nuts_seeds, grains and pork_poultry (7gm/day,35gm/day,28gm/day) causes high impact on cropland as compared to other items of diet. Eggs and fish (6gm/day and 1gm/day) have less average values that impact on cropland. Each violin represents the density of data points for a specific category of "disagg.value.x." The width of the violin at different y-values indicates the density of data points at that level of cropland impact. The mean is indicated by a white dot within each violin.

Q ~ Q Association

environmental_impacts_freshwater_use ~ environmental_impacts_land_use

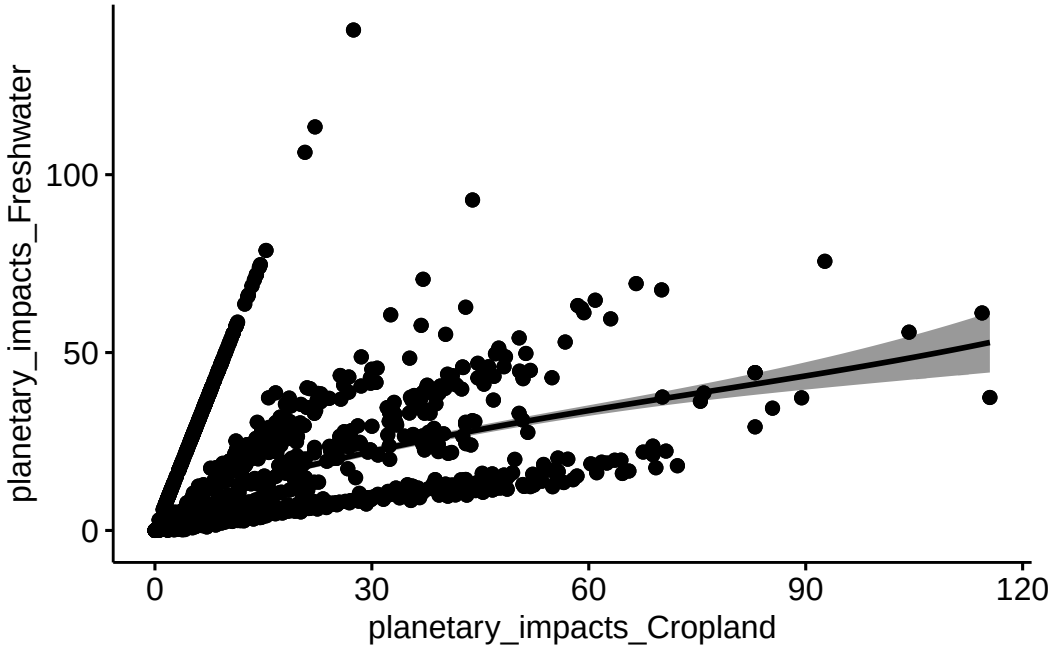
"planetary_impacts_Cropland" variable represents environmental impact of using freshwater (use in %) on diet and nutrition in certain regions.

"planetary_impacts_Freshwater" variable represents environmental impact of land (use in %) on diet and nutrition in certain regions.

```
cor(mydata$planetary_impacts_Cropland,mydata$planetary_impacts_Freshwater,
    use="pairwise.complete.obs")
```

```
[1] 0.5549602
```

```
ggscatter(mydata,
  x="planetary_impacts_Cropland", y = "planetary_impacts_Freshwater",
  add = "loess", conf.int = TRUE)
```



The scatter plot you provided shows the relationship between "planetary_impacts_Cropland" (on the horizontal axis in %) and "planetary_impacts_Freshwater" (on the vertical axis in %) where "planetary_impacts_Cropland" has a strong, positive and totally linear correlation with "planetary_impact_Freshwater", r = 0.55. This suggests that as the impact on cropland increases, the impact on freshwater also tends to increase.

B ~ Q Association

disagg.value.x ~ planetary_impacts_Freshwater

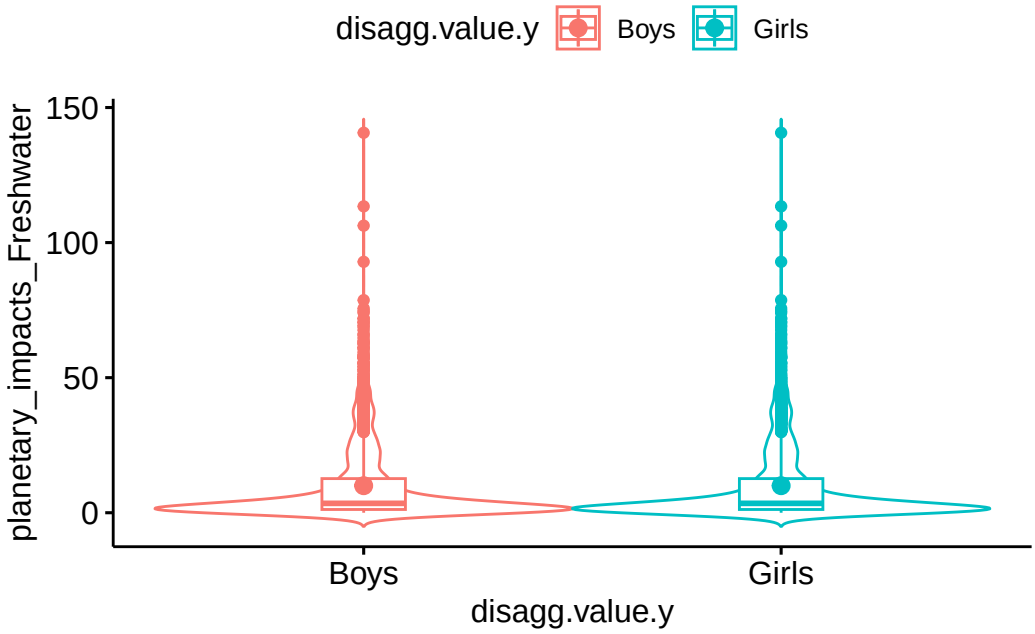
"*disagg.value.x*" variable represents diet items(fruits,vegetables,beef,etc) grams/day in particular country.

"planetary_impacts_Freshwater" variable represents environmental impact of using freshwater (use in %) on diet and nutrition in certain countries.

```
mydata$disagg.value.y <- as.factor(mydata$disagg.value.y)
mydata %>%
group_by(disagg.value.y) %>%
summarize(
N = n(),
Mean = mean(planetary_impacts_Freshwater, na.rm = TRUE),
Median = median(planetary_impacts_Freshwater, na.rm = TRUE),
SD = sd(planetary_impacts_Freshwater, na.rm = TRUE),
IQR = IQR(planetary_impacts_Freshwater, na.rm = TRUE)
)
```

```
# A tibble: 2 x 6
  disagg.value.y     N  Mean Median    SD   IQR
  <fct>          <int> <dbl>  <dbl> <dbl> <dbl>
1 Boys            1978  10.0   3.47  14.6  11.4
2 Girls           1978  10.0   3.47  14.6  11.4
```

```
ggviolin(mydata,
  x="disagg.value.y", y = "planetary_impacts_Freshwater",
  color = "disagg.value.y", add = c("mean", "boxplot"))
```



The violin plot you provided compares the distribution of the variable "planetary_impacts_Freshwater" across two groups, "Boys" and "Girls," represented by the variable "disagg.value.y". The width of each violin at different y-values indicates the density of data points. Wider sections represent higher density, showing where most data points lie. Both the values have same statisitics having equal number of frequency(1978), mean(10), median(3.47), standard deviation(14.56) and IQR(11.43) also same for both values in "disagg.values.y".