NS125 PCW Session 8

2022-09-26

Using packages and inspecting data

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(readxl)
library(ellipse)
```

```
##
## Attaching package: 'ellipse'

## The following object is masked from 'package:graphics':
##
##     pairs
```

```
df <- read_excel("./session-07/data.xlsx", skip = 1, col_names = TRUE, na="NIL")
head(df)
```

```
## # A tibble: 6 x 43
##      WP5 Country     ISO   Region Sub_r~1 Sampl~2 Sampl~3 Sampl~4 Popul~5  WPCIAS
##    <dbl> <chr>       <chr> <chr>  <chr>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1     70 Afghanistan AFG   Asia   Wester~       0    1010    1010  1.37e7 6.43
## 2     56 Angola      AGO   Africa Centra~       0    1000    1000  9.55e6 0.563
## 3     87 Argentina   ARG   Latin~ Southe~    1000    1000    2000  2.96e7 9.07
## 4     88 Armenia     ARM   Europe Easter~    1000    1000    2000  2.36e6 0.00485
## 5     47 Australia   AUS   Europ~ Wester~       0    1005    1005  1.73e7 6.98
## 6     89 Austria     AUT   Europe Wester~       0    1001    1001  7.07e6 0.012
## # ... with 33 more variables: SPCIAS <dbl>, APCIAS <dbl>, HDI <dbl>,
## #   CO2emi <dbl>, GDPperUS <dbl>, glo_tot <dbl>, Efcon <dbl>, TotBioCap <dbl>,
## #   VA <dbl>, PS <dbl>, GovE <dbl>, RQ <dbl>, RL <dbl>, CC <dbl>, WGI <dbl>,
## #   Aware <dbl>, Unaware <dbl>, RF_aware <dbl>, AUC_aware <dbl>,
## #   CI_aware <dbl>, Top_aware <chr>, Top2_aware <chr>, Top3_aware <chr>,
## #   Ratio_aware <chr>, Serious <dbl>, Not_serious <dbl>, RF_serious <dbl>,
## #   AUC_serious <dbl>, CITree_serious <dbl>, Top1_serious <chr>, ...
```

```
unique(df$Region)
```

```
## [1] "Asia"                 "Africa"
```

```
## [3] "Latin America & Caribbean" "Europe"
## [5] "Europe (Oceania)"          "Europe (North America)"
```

```
unique(df$Sub_region)
```

```
##  [1] "Western Asia"              "Central Africa"
##  [3] "Southern South America"    "Eastern Europe"
##  [5] "Western Europe (AustraliaNZ)"  "Western Europe"
##  [7] "Coastal West Africa"       "Sahelian Africa"
##  [9] "Southern Asia"             "Central America"
## [11] "Andean South America"      "Northern South America"
## [13] "Southern Africa"           "Western Europe (North America)"
## [15] "Northeast Asia"            "East Africa"
## [17] "Caribbean Islands"         "North Africa"
## [19] "Southeast Asia"            "Middle East"
```

```
nrow(df[df$Region == 'Europe (North America)', ])
```

```
## [1] 2
```

```
nrow(df[df$Region == 'Europe (Oceania)', ])
```

```
## [1] 1
```

Factorizing columns

```
factor_cols <- c("Top_aware", "Top2_aware", "Top3_aware", "Top1_serious", "Top2_serious", "Top3_serious"
df[,factor_cols] <- lapply(df[,factor_cols], factor)
head(df)
```

```
## # A tibble: 6 x 43
##      WP5 Country     ISO   Region Sub_r~1 Sampl~2 Sampl~3 Sampl~4 Popul~5  WPCIAS
##    <dbl> <chr>       <chr> <fct>  <fct>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1     70 Afghanistan AFG   Asia   Wester~       0    1010    1010  1.37e7 6.43
## 2     56 Angola      AGO   Africa Centra~       0    1000    1000  9.55e6 0.563
## 3     87 Argentina   ARG   Latin~ Southe~    1000    1000    2000  2.96e7 9.07
## 4     88 Armenia     ARM   Europe Easter~    1000    1000    2000  2.36e6 0.00485
## 5     47 Australia   AUS   Europ~ Wester~       0    1005    1005  1.73e7 6.98
## 6     89 Austria     AUT   Europe Wester~       0    1001    1001  7.07e6 0.012
## # ... with 33 more variables: SPCIAS <dbl>, APCIAS <dbl>, HDI <dbl>,
## #   CO2emi <dbl>, GDPperUS <dbl>, glo_tot <dbl>, Efcon <dbl>, TotBioCap <dbl>,
## #   VA <dbl>, PS <dbl>, GovE <dbl>, RQ <dbl>, RL <dbl>, CC <dbl>, WGI <dbl>,
## #   Aware <dbl>, Unaware <dbl>, RF_aware <dbl>, AUC_aware <dbl>,
## #   CI_aware <dbl>, Top_aware <fct>, Top2_aware <fct>, Top3_aware <fct>,
## #   Ratio_aware <chr>, Serious <dbl>, Not_serious <dbl>, RF_serious <dbl>,
## #   AUC_serious <dbl>, CITree_serious <dbl>, Top1_serious <fct>, ...
```
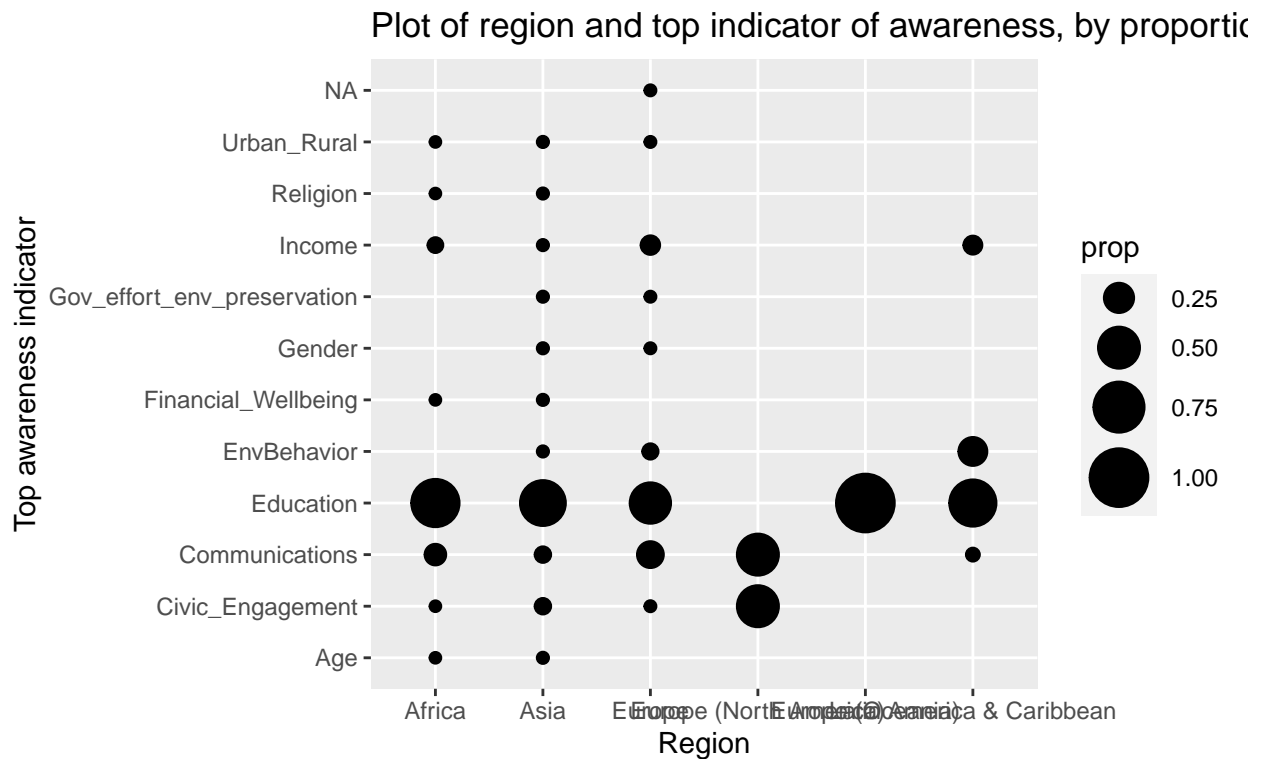
Plots

```
ggplot(df, aes(x = Region, y = Top_aware)) +
  geom_count(aes(color = ..n.., size = ..n..)) +
  scale_size_area(max_size = 10) +
  guides(color = 'legend') +
  labs(title="Plot of region and top indicator of awareness, by count", y = "Top awareness indicator",
       caption="Top indicator of awareness by regions.\nEducation is the best indicator in all regions
```

Plot of region and top indicator of awareness, by count
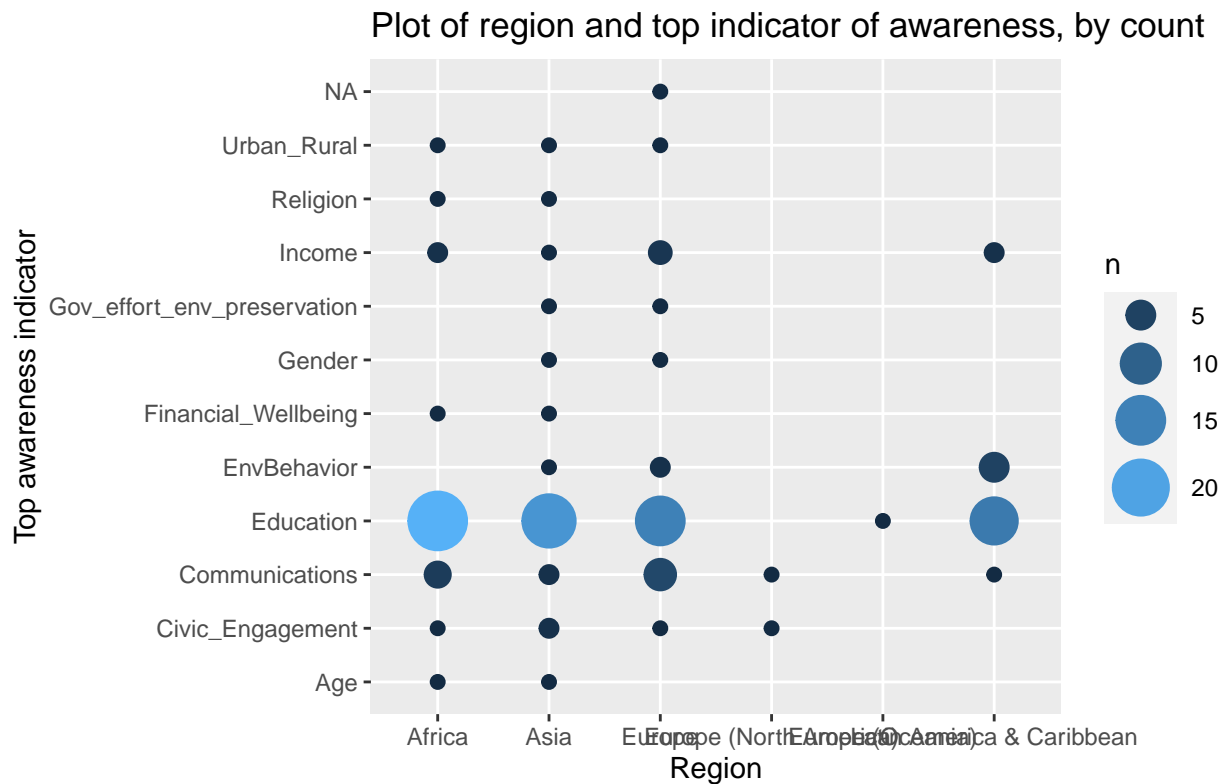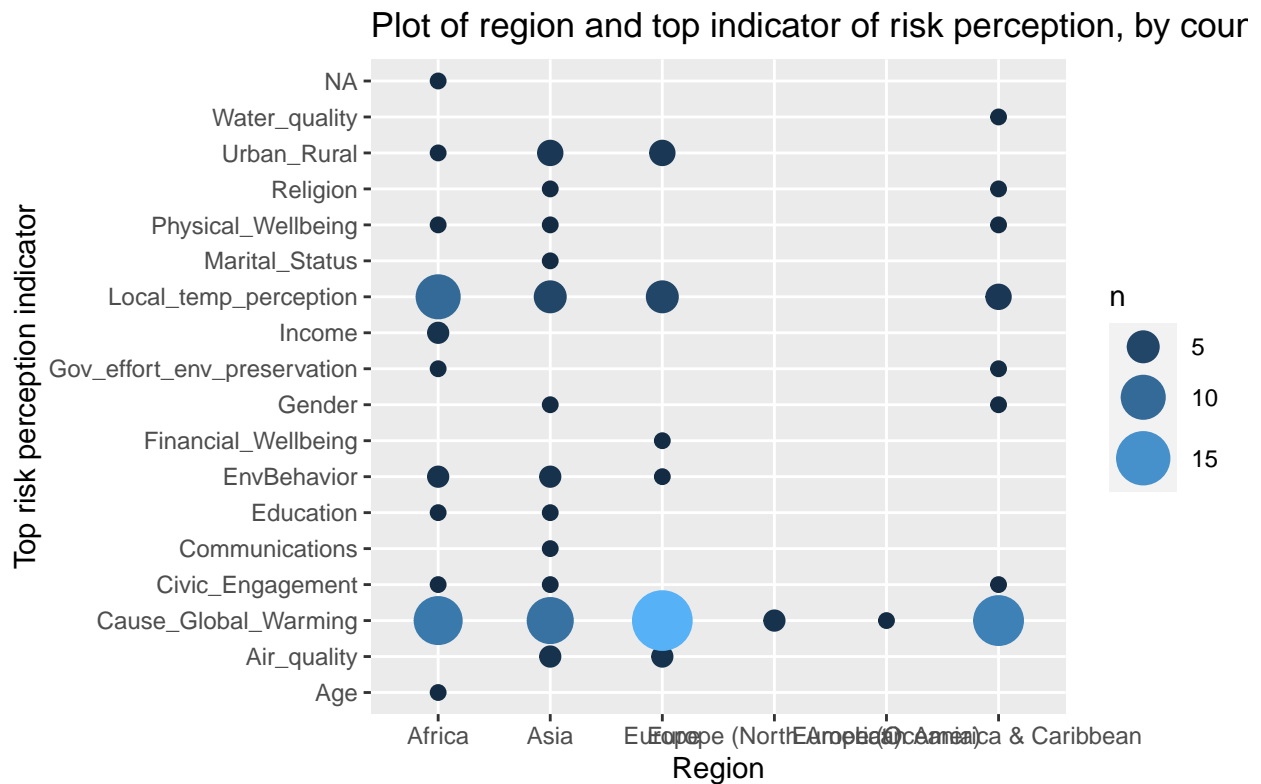

Top indicator of awareness by regions.
Education is the best indicator in all regions except North America.

```
ggplot(df, aes(x = Region, y = Top_aware)) +
  geom_count(aes(size = after_stat(prop),  group = Region)) +
  scale_size_area(max_size = 10) +
  labs(title="Plot of region and top indicator of awareness, by proportion", y = "Top awareness indicat
  caption="Top indicator of awareness by regions, by proportion.\nThis is to account for the fact that
  theme(plot.caption = element_text(hjust = 0))
```

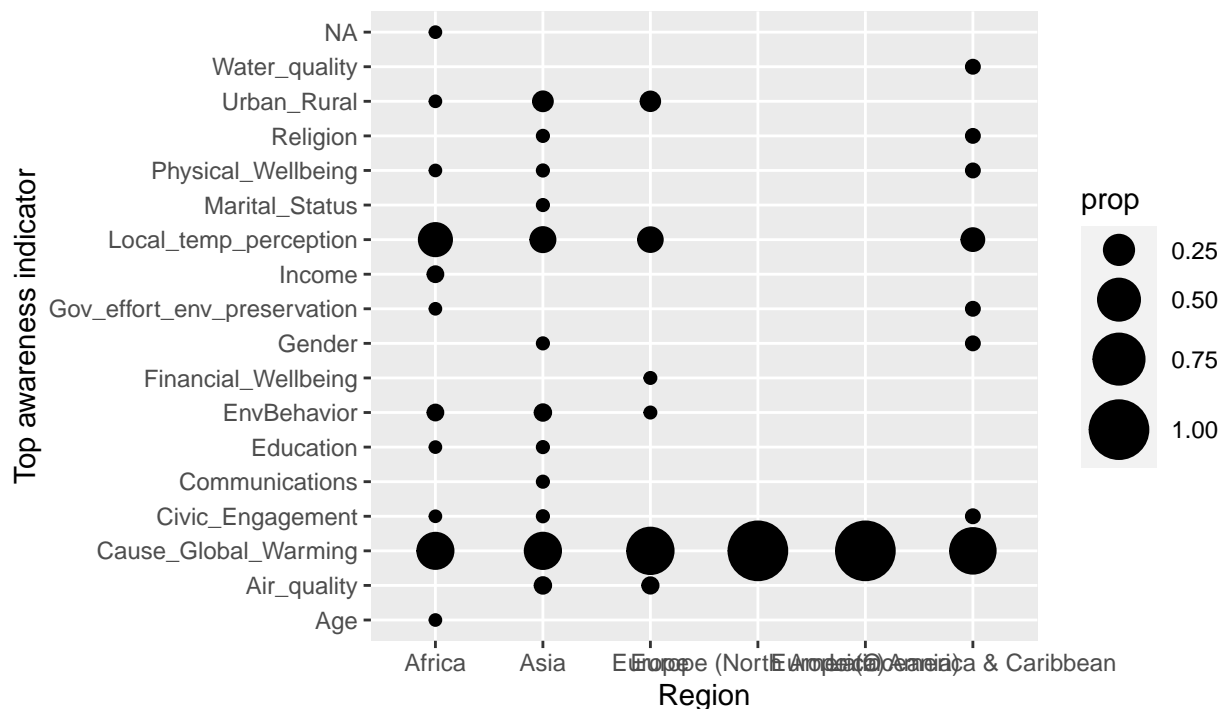## Plot of region and top indicator of awareness, by proportion

Top indicator of awareness by regions, by proportion.
This is to account for the fact that North America only has 2 countries and Oceania
Education is the best indicator in all regions except North America.

```
ggplot(df, aes(x = Region, y = Top1_serious)) +
  geom_count(aes(color = ..n.., size = ..n..)) +
  scale_size_area(max_size = 10) +
  guides(color = 'legend') +
  labs(title="Plot of region and top indicator of risk perception, by count",
       y = "Top risk perception indicator",
       caption="Top indicator of awareness by regions.\nUnderstanding of global warming cause is the bes
  theme(plot.caption = element_text(hjust = 0))
```

Plot of region and top indicator of risk perception, by cour

Top indicator of awareness by regions.
Understanding of global warming cause is the best indicator in all regions.

```
ggplot(df, aes(x = Region, y = Top1_serious)) +
  geom_count(aes(size = after_stat(prop),  group = Region)) +
  scale_size_area(max_size = 10) +
  labs(title="Plot of region and top indicator of risk perception, by proportion", y = "Top awareness in
       caption="Top indicator of awareness by region, by proportion.\nThis is to account for the fact th
  scale_color_brewer(palette="Blues") +
  theme(plot.caption = element_text(hjust = 0))
```

Plot of region and top indicator of risk perception, by prop

Top indicator of awareness by region, by proportion.
This is to account for the fact that North America only has 2 countries and Oceania
Understanding of global warming cause is the best indicator in all regions.

```
#Group global south and global north
Global_north <- c("Europe", "Europe (Oceania)", "Europe (North America)")
df$global_south_north = ifelse(df$Region %in% Global_north, "North", "South")
df$global_south_north <- as.factor(df$global_south_north)
head(df)
```

```
## # A tibble: 6 x 44
##      WP5 Country    ISO   Region Sub_r~1 Sampl~2 Sampl~3 Sampl~4 Popul~5  WPCIAS
##    <dbl> <chr>      <chr> <fct>  <fct>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1     70 Afghanistan AFG  Asia   Wester~       0    1010    1010  1.37e7 6.43
## 2     56 Angola     AGO   Africa Centra~       0    1000    1000  9.55e6 0.563
## 3     87 Argentina  ARG   Latin~ Southe~    1000    1000    2000  2.96e7 9.07
## 4     88 Armenia    ARM   Europe Easter~    1000    1000    2000  2.36e6 0.00485
## 5     47 Australia  AUS   Europ~ Wester~       0    1005    1005  1.73e7 6.98
## 6     89 Austria    AUT   Europe Wester~       0    1001    1001  7.07e6 0.012
## # ... with 34 more variables: SPCIAS <dbl>, APCIAS <dbl>, HDI <dbl>,
## #   CO2emi <dbl>, GDPperUS <dbl>, glo_tot <dbl>, Efcon <dbl>, TotBioCap <dbl>,
## #   VA <dbl>, PS <dbl>, GovE <dbl>, RQ <dbl>, RL <dbl>, CC <dbl>, WGI <dbl>,
## #   Aware <dbl>, Unaware <dbl>, RF_aware <dbl>, AUC_aware <dbl>,
## #   CI_aware <dbl>, Top_aware <fct>, Top2_aware <fct>, Top3_aware <fct>,
## #   Ratio_aware <chr>, Serious <dbl>, Not_serious <dbl>, RF_serious <dbl>,
## #   AUC_serious <dbl>, CITree_serious <dbl>, Top1_serious <fct>, ...
```

```
#Get average awareness
awareness <- df %>%
  group_by(global_south_north) %>%
  summarise_at(vars("Aware", "Serious"), mean)
```

```
awareness
```

```
## # A tibble: 2 x 3
##   global_south_north Aware Serious
##   <fct>              <dbl>   <dbl>
## 1 North               89.0    68.2
## 2 South               53.5    83.1
```

```
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
awareness <- as.data.table(awareness)
awareness_df <- melt(awareness, id.vars = "global_south_north", variable.name = "category",
        value.name="percentage")
head(awareness_df)
```

```
##    global_south_north category percentage
## 1:              North    Aware   89.03246
## 2:              South    Aware   53.49043
## 3:              North   Serious   68.16506
## 4:              South   Serious   83.14310
```

```
ggplot(awareness_df,aes(x = global_south_north, y=percentage, fill=category, color=category)) +
    geom_bar(position = "dodge", stat = "identity") +
  labs(title="Awareness and risk perception percentage in global South vs North", y = "Percentage", x =
        caption="Awareness in global North (89%) is higher than global South (53%).\nHowever, of those wh
  theme(plot.caption = element_text(hjust = 0))
```

Awareness and risk perception percentage in global South vs North
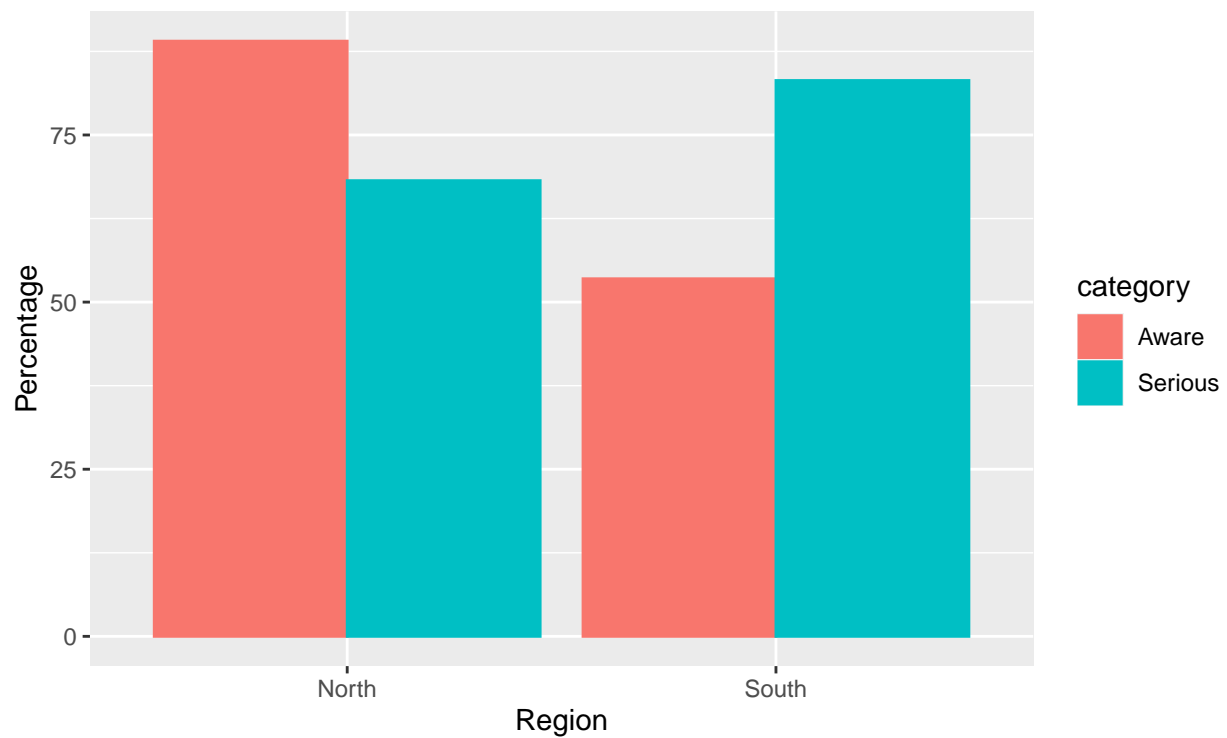
Awareness in global North (89%) is higher than global South (53%).
However, of those who are aware, global South take climate change risks more seriously (83%) than global N