R Visualization 2

Tangkwa Siwanuch S

2025-11-07

```r
# Install and load library
#install.packages("tidyverse")
#install.packages("ggrepel")

library(tidyverse)
library(ggrepel)

# Read data from csv file
netusage <- read_csv("InternetUsage.csv")

# Data Overview
glimpse(netusage)
```

```
## Rows: 3,500
## Columns: 24
## $ id                      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14~
## $ age                     <dbl> 40, 27, 31, 41, 26, 37, 18, 33, 43, 41, 47, 1~
## $ gender                  <chr> "Female", "Male", "Male", "Female", "Female",~
## $ region                  <chr> "Asia", "Africa", "North America", "Middle Ea~
## $ income_level            <chr> "High", "Lower-Mid", "Lower-Mid", "Low", "Low~
## $ education_level         <chr> "High School", "Master", "Bachelor", "Master"~
## $ daily_role              <chr> "Part-time/Shift", "Full-time Employee", "Ful~
## $ device_hours_per_day    <dbl> 3.54, 5.65, 8.87, 4.05, 13.07, 4.59, 4.95, 6.~
## $ phone_unlocks           <dbl> 45, 100, 181, 94, 199, 73, 119, 82, 155, 38, ~
## $ notifications_per_day   <dbl> 561, 393, 231, 268, 91, 198, 553, 184, 309, 1~
## $ social_media_mins       <dbl> 98, 174, 595, 18, 147, 9, 61, 48, 16, 249, 31~
## $ study_mins              <dbl> 34, 102, 140, 121, 60, 85, 188, 155, 116, 155~
## $ physical_activity_days  <dbl> 7, 2, 1, 4, 1, 0, 4, 3, 4, 5, 4, 7, 4, 4, 1, ~
## $ sleep_hours             <dbl> 9.123800, 8.837517, 6.486743, 7.600504, 5.197~
## $ sleep_quality           <dbl> 3.353627, 2.908147, 2.889213, 3.097488, 2.786~
## $ anxiety_score           <dbl> 9.926651, 4.000000, 4.000000, 7.093357, 7.028~
## $ depression_score        <dbl> 5, 4, 8, 9, 15, 4, 1, 8, 18, 0, 12, 2, 9, 0, ~
## $ stress_level            <dbl> 6.593289, 4.126926, 1.429139, 4.995512, 9.448~
## $ happiness_score         <dbl> 8.000000, 8.100000, 7.600000, 7.800000, 4.200~
## $ focus_score             <dbl> 23, 35, 15, 28, 70, 64, 15, 70, 53, 73, 59, 2~
## $ high_risk_flag          <dbl> 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ device_type             <chr> "Android", "Laptop", "Android", "Tablet", "An~
## $ productivity_score      <dbl> 70.0000, 64.0000, 65.2993, 80.0000, 65.2993, ~
## $ digital_dependence_score <dbl> 25.70000, 30.10000, 40.60000, 36.68415, 48.40~
```

```r
str(netusage)
```

```
## spc_tbl_ [3,500 x 24] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
```

```
##  $ id                    : num [1:3500] 1 2 3 4 5 6 7 8 9 10 ...
##  $ age                   : num [1:3500] 40 27 31 41 26 37 18 33 43 41 ...
##  $ gender                : chr [1:3500] "Female" "Male" "Male" "Female" ...
##  $ region                : chr [1:3500] "Asia" "Africa" "North America" "Middle East" ...
##  $ income_level          : chr [1:3500] "High" "Lower-Mid" "Lower-Mid" "Low" ...
##  $ education_level       : chr [1:3500] "High School" "Master" "Bachelor" "Master" ...
##  $ daily_role            : chr [1:3500] "Part-time/Shift" "Full-time Employee" "Full-time Employee"
##  $ device_hours_per_day  : num [1:3500] 3.54 5.65 8.87 4.05 13.07 ...
##  $ phone_unlocks         : num [1:3500] 45 100 181 94 199 73 119 82 155 38 ...
##  $ notifications_per_day : num [1:3500] 561 393 231 268 91 198 553 184 309 110 ...
##  $ social_media_mins     : num [1:3500] 98 174 595 18 147 9 61 48 16 249 ...
##  $ study_mins            : num [1:3500] 34 102 140 121 60 85 188 155 116 155 ...
##  $ physical_activity_days: num [1:3500] 7 2 1 4 1 0 4 3 4 5 ...
##  $ sleep_hours           : num [1:3500] 9.12 8.84 6.49 7.6 5.2 ...
##  $ sleep_quality         : num [1:3500] 3.35 2.91 2.89 3.1 2.79 ...
##  $ anxiety_score         : num [1:3500] 9.93 4 4 7.09 7.03 ...
##  $ depression_score      : num [1:3500] 5 4 8 9 15 4 1 8 18 0 ...
##  $ stress_level          : num [1:3500] 6.59 4.13 1.43 5 9.45 ...
##  $ happiness_score       : num [1:3500] 8 8.1 7.6 7.8 4.2 10 7.7 8.6 8.3 9.2 ...
##  $ focus_score           : num [1:3500] 23 35 15 28 70 64 15 70 53 73 ...
##  $ high_risk_flag        : num [1:3500] 0 0 0 1 1 0 0 0 0 0 ...
##  $ device_type           : chr [1:3500] "Android" "Laptop" "Android" "Tablet" ...
##  $ productivity_score    : num [1:3500] 70 64 65.3 80 65.3 ...
##  $ digital_dependence_score: num [1:3500] 25.7 30.1 40.6 36.7 48.4 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   id = col_double(),
##   ..   age = col_double(),
##   ..   gender = col_character(),
##   ..   region = col_character(),
##   ..   income_level = col_character(),
##   ..   education_level = col_character(),
##   ..   daily_role = col_character(),
##   ..   device_hours_per_day = col_double(),
##   ..   phone_unlocks = col_double(),
##   ..   notifications_per_day = col_double(),
##   ..   social_media_mins = col_double(),
##   ..   study_mins = col_double(),
##   ..   physical_activity_days = col_double(),
##   ..   sleep_hours = col_double(),
##   ..   sleep_quality = col_double(),
##   ..   anxiety_score = col_double(),
##   ..   depression_score = col_double(),
##   ..   stress_level = col_double(),
##   ..   happiness_score = col_double(),
##   ..   focus_score = col_double(),
##   ..   high_risk_flag = col_double(),
##   ..   device_type = col_character(),
##   ..   productivity_score = col_double(),
##   ..   digital_dependence_score = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```r
## Visualization No. 1

# Distinct value of daily_role
netusage %>%
  count(daily_role)
```

```
## # A tibble: 5 x 2
##   daily_role            n
##   <chr>             <int>
## 1 Caregiver/Home      149
## 2 Full-time Employee 1429
## 3 Part-time/Shift     659
## 4 Student             905
## 5 Unemployed_Looking  358
```

```r
# Transform daily_role to shorter name
netusage %>%
  mutate(daily_role_short = case_when(
    daily_role == "Caregiver/Home" ~ "Caregiver",
    daily_role == "Full-time Employee" ~ "Full-time",
    daily_role == "Part-time/Shift" ~ "Part-time",
    daily_role == "Unemployed_Looking" ~ "Unemployed",
    TRUE ~ daily_role
  )) %>%

# Use ggplot: scatter plots of 2 continuous variables, multiple plots by 2 discrete variables

ggplot(data = . ,
       mapping = aes(x=digital_dependence_score, y=anxiety_score)) +
  geom_point(size=1.5, alpha=0.8) +
  geom_smooth() +
  facet_grid(daily_role_short ~ region) +
  theme_minimal() +
  theme(
  strip.text.y = element_text(size = 6)) +
  labs(title = "Digital Dependence vs. Anxiety Level by Employment Type",
       x = "Digital Dependence Score",
       y = "Anxiety Score")
```

Digital Dependence vs. Anxiety Level by Employment Type
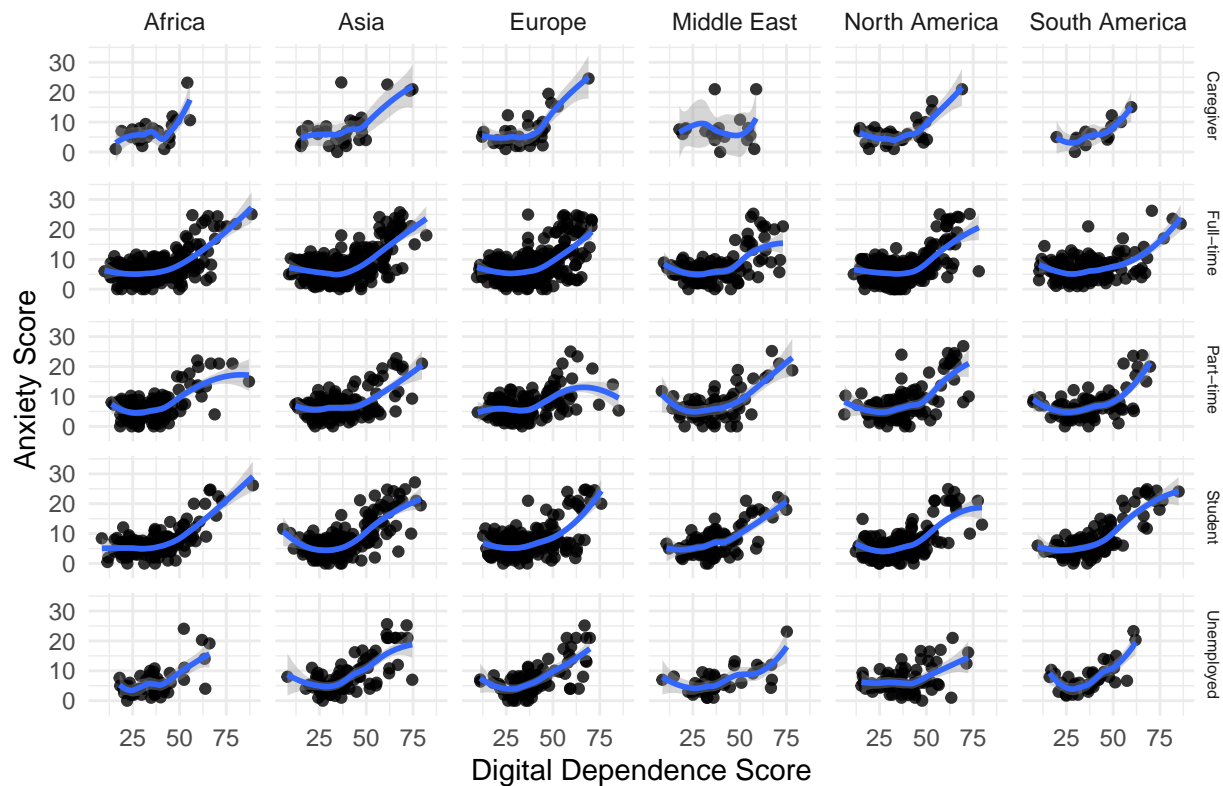

```
## Visualization No. 2

# Transform age column from continuous to discrete
netusage %>%
  mutate(age_range = case_when(
    age<= 10  ~ "1_10",
    age<=20 ~ "11_20",
    age<=30 ~ "21_30",
    age<=40 ~ "31_40",
    age<=50 ~ "41_50",
    TRUE ~ "51_more"
  )) %>%
  mutate(region_2 = case_when(
    region == "North America"  ~ "North\nAmerica",
    region == "South America"  ~ "South\nAmerica",
    region == "Middle East"  ~ "Middle\nEast",
    TRUE ~ region
  )) %>%
# Use ggplot: box plots of 1 continuous variable by 2 discrete variables
ggplot(data = .,
       mapping = aes(x=region_2, y=digital_dependence_score)) +
  geom_boxplot() +
  facet_wrap(~age_range) +
  theme_minimal() +
  theme(
  strip.text.y = element_text(size = 6)) +
  labs(title = "Digital Dependence vs. Age By Region",
```

```
x = "Age",
y = "Digital Dependence Score")
```

Digital Dependence vs. Age By Region