

STAT 445/645 Assignment Cover Page

Student Name

SFU Student Number

SFU email address

Assignment Number

Due Date

Provide references for any data sets used in this assignment

List software used in this assignment.

List **ALL** resources used to complete this assignment, including books, internet sources and people.

☐ I personally completed the computations and wrote the solutions submitted in this document.

A1_Q1

Joohyeok

2024-02-05

```
library(ggplot2)
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.3      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ lubridate  1.9.3      ✓ tibble     3.2.1
## ✓ purrr      1.0.2      ✓ tidyr      1.3.0
## — Conflicts — tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
come errors
```

```
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2
```

```
library(readxl)
library(car)
```

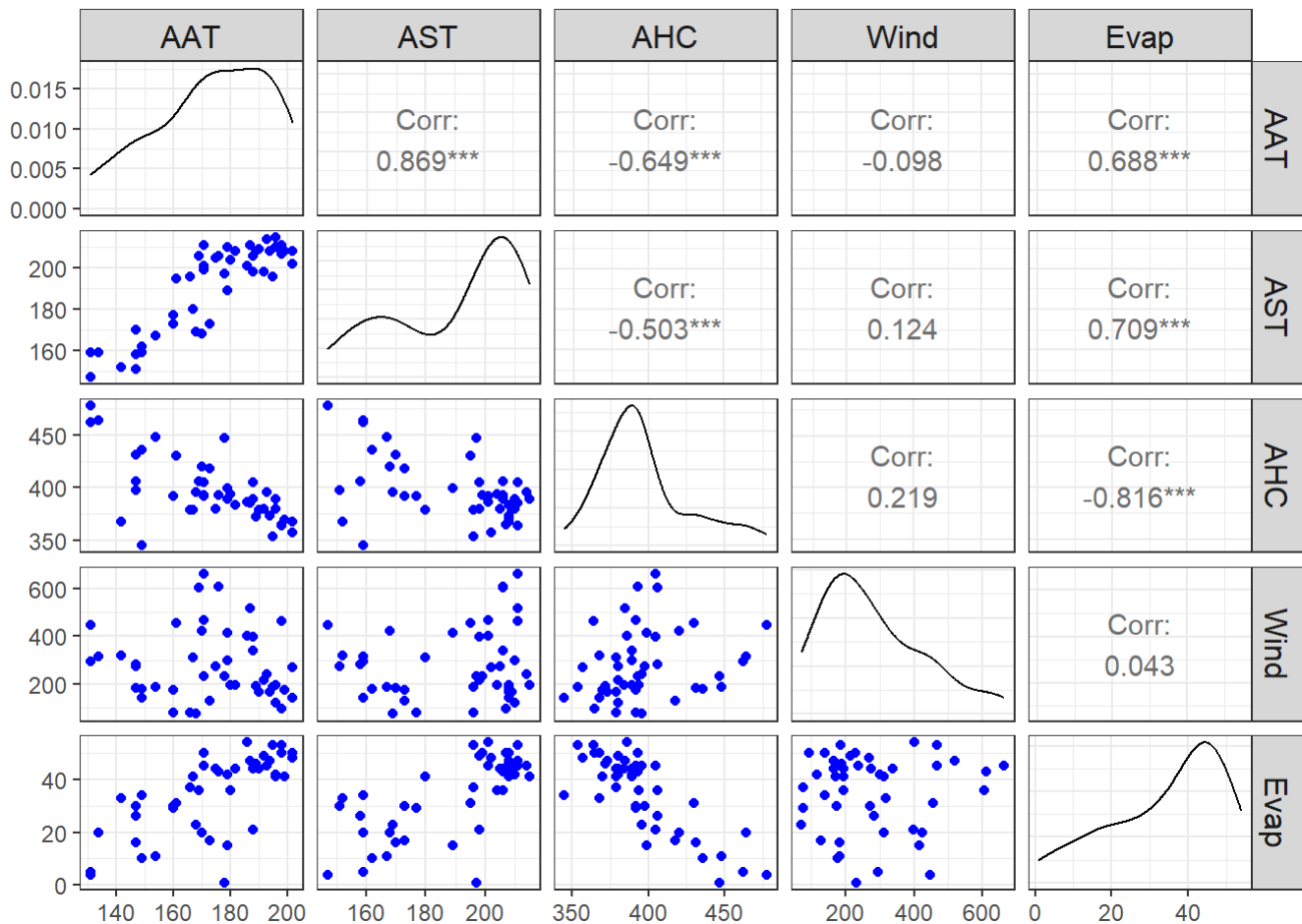
```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##   recode
##
## The following object is masked from 'package:purrr':
##
##   some
```

```
df<-read_excel("cutdown_environment.xlsx")
```

a. Create a matrix scatter plot for the data.

```
ggpairs(df,columns=c("AAT","AST","AHC","Wind","Evap"),
        lower=list(continuous = wrap("points", color = "blue")))+
  theme_bw()+
  theme(strip.text=element_text(size=12))
```

```
## Warning in warn_if_args_exist(list(...)): Extra arguments: "columns" are being
## ignored. If these are meant to be aesthetics, submit them using the 'mapping'
## variable within ggpairs with ggplot2::aes or ggplot2::aes_string.
```



- b. The correlation between AAT and AST is 0.869. It indicates a strong positive linear relationship which is seen in second row first column plot. Since this value is close to 1, there is a strong dependency.
- c. The correlation between EVAP and AHC is -0.816. It indicates a strong negative linear relationship which is seen in fifth row third column plot. Since this value is close to -1, there is a strong dependency.
- d. The correlation between AAT and WIND is -0.098. It indicates a non-linear relationship which is seen in fourth row first column plot. Since this value is close to 0, there is no dependency.

Q2_Final

Joohyeok

2024-02-09

```
library(ggplot2)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2
```

```
library(readr)

data=read.table("baker_corn_field_data.txt")

colnames(data) <- LETTERS[1:ncol(data)]

corn<-data[,c("B","C","G","H")]

Color<-as.factor(corn$B)

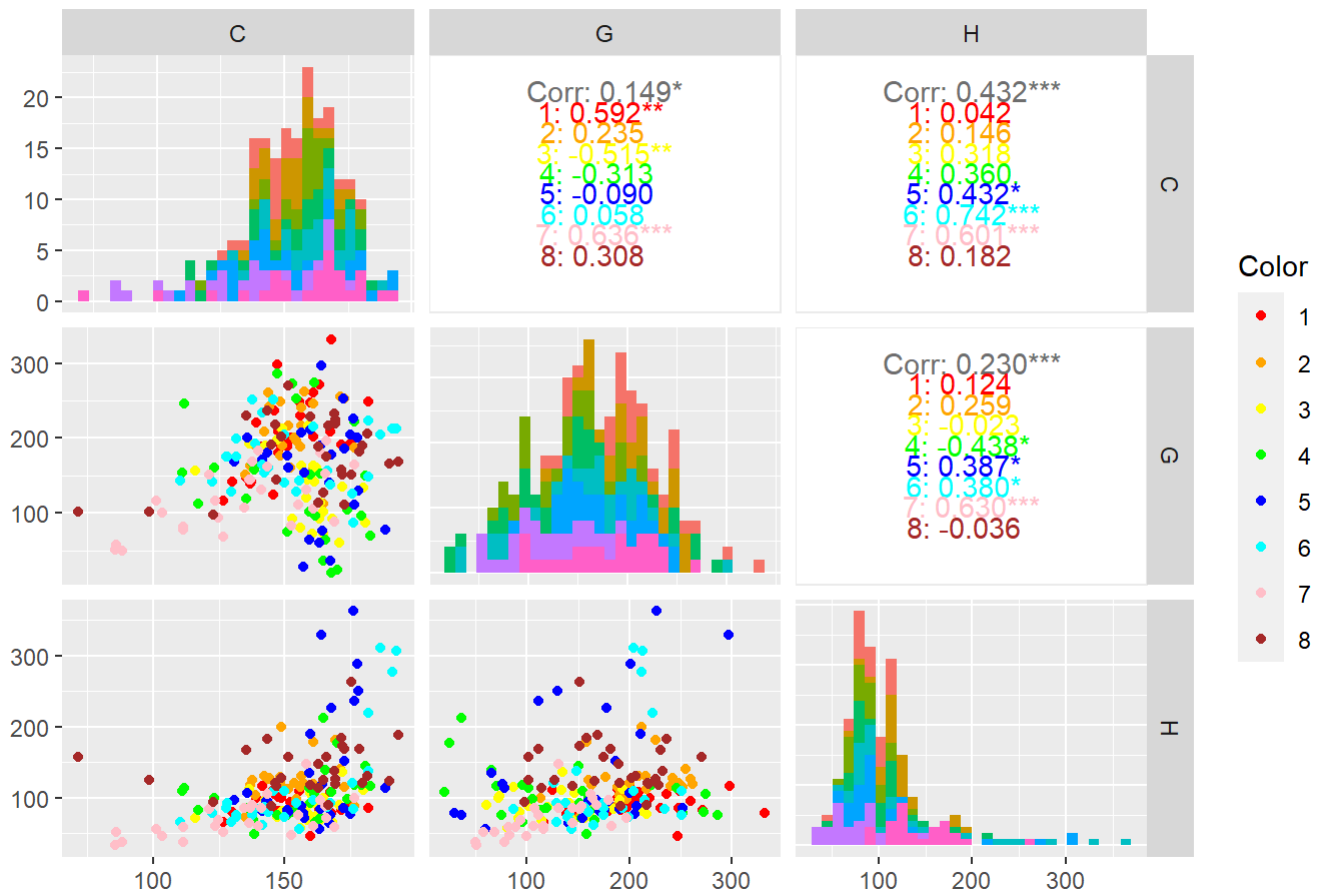
colors <- c("red", "orange", "yellow", "green", "blue", "cyan", "pink", "brown")
```

a. create a 3x3 matrix scatter plot

```
ggpairs(corn[,c("C","G","H")],mapping = aes(color = Color),legend=4,
        diag = list(continuous = "barDiag"))+
  ggtitle("Matrix Scatter Plot for Corn Field Data")+
  theme(plot.title=element_text(hjust=0.5))+
  scale_color_manual(values = colors)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Matrix Scatter Plot for Corn Field Data



b. List groups that appear to be relatively more tightly clustered than other groups

- A correlation value close to 1 or -1 indicates a strong clustering of points
- C vs G plot is 2 row 1 column. Group 1, 3 and 7 looks clustered than other groups
- C vs H plot is 3 row 1 column. Group 6 and 7 looks clustered than other groups
- G vs H plot is 3 row 2 column. Group 7 looks clustered than other groups.

c. List groups that appear to be relatively widely dispersed

- A correlation value close to 0 indicates that the points are widely dispersed
- C vs G plot is 2 row 1 column. Group 5 and 6 looks widely dispersed
- C vs H plot is 3 row 1 column. Group 1 looks widely dispersed
- G vs H plot is 3 row 2 column. Group 3 and 8 looks widely dispersed