Untitled

```r
library(tidyverse)
```

```
-- Attaching core tidyverse packages ---------------------- tidyverse 2.0.0 --
v dplyr     1.1.2     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```r
library(ggplot2)
library(gt)
```

```r
data <- read.csv("../data/cleaned_data/data_cleaned.csv")
```

```r
data %>%
  mutate(Decision = fct_relevel(Decision, "Admit", "Waitlist", "Decline")) %>%
  summarize("Test Score" = mean(TestScore), "Writing Score" = mean(WritingScore), "GPA" =
  gt() %>%
  fmt_number(
    columns = c(`GPA`),
    decimals = 2,
    use_seps = FALSE
  ) %>%
  fmt_number(
    columns = c(`Test Score`, `Writing Score`),
```

```
    decimals = 1,
    use_seps = FALSE
  ) %>%
  tab_header(title = md("Average Scores for all students"))
```

Average Scores for all students

Test Score	Writing Score	GPA
868.9	82.6	3.49

```
data %>%
  mutate(Decision = fct_relevel(Decision, "Admit", "Waitlist", "Decline")) %>%
  group_by(Decision) %>%
  summarize("Test Score" = mean(TestScore), "Writing Score" = mean(WritingScore), "GPA" =
  gt() %>%
  fmt_number(
    columns = c(`GPA`),
    decimals = 2,
    use_seps = FALSE
  ) %>%
  fmt_number(
    columns = c(`Test Score`, `Writing Score`),
    decimals = 1,
    use_seps = FALSE
  )%>%
  tab_header(title = md("Average Scores for students by decision"))
```
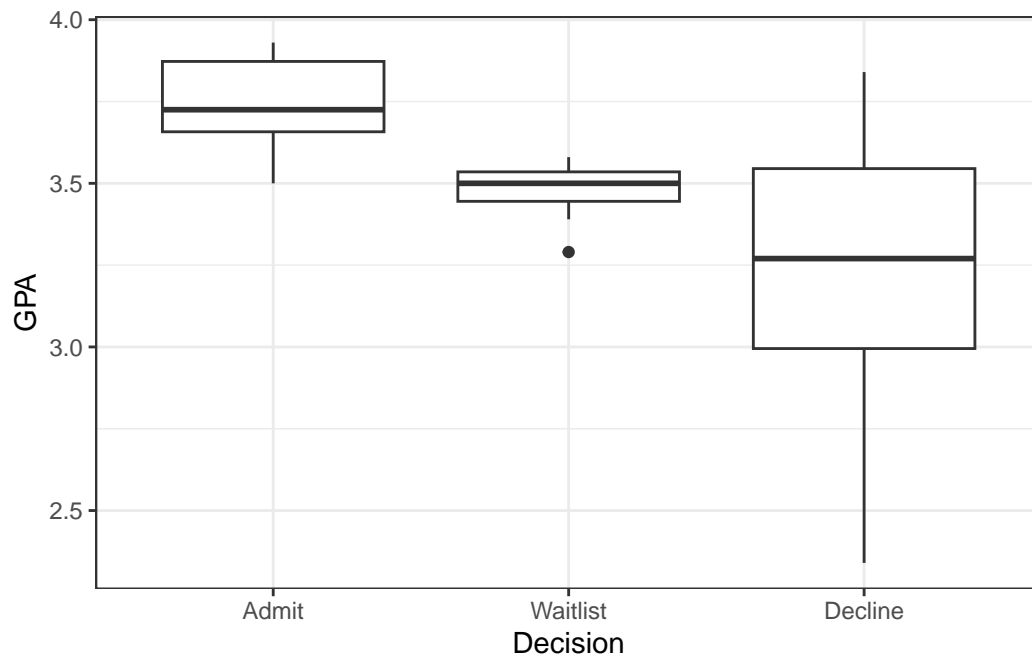
Average Scores for students by decision

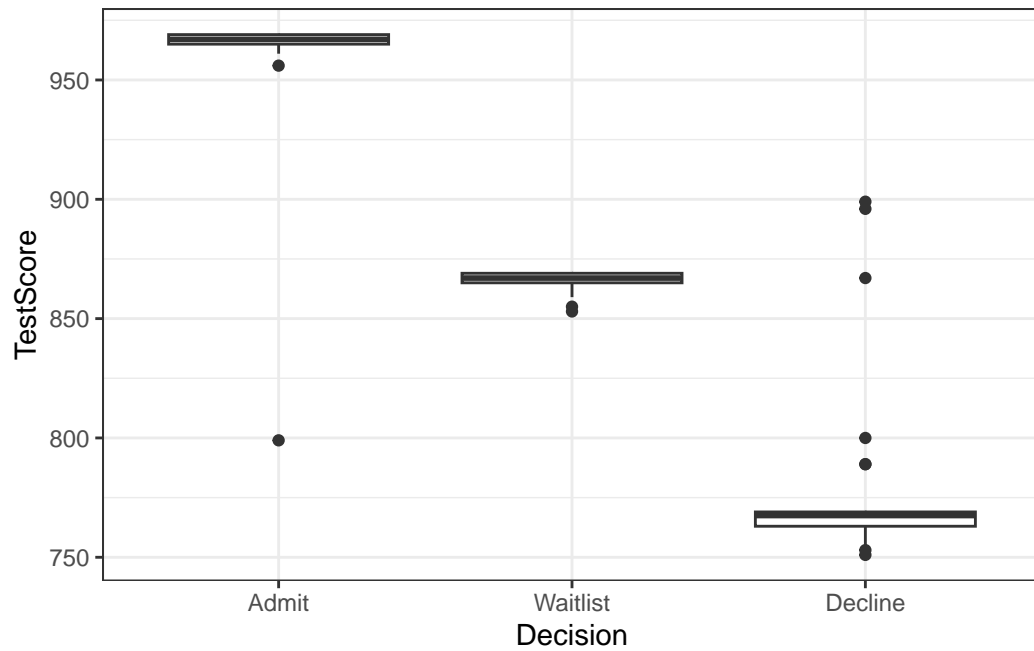| Decision	Test Score	Writing Score	GPA
Admit	960.2	91.2	3.74
Waitlist	865.7	82.5	3.49
Decline	780.1	74.1	3.25

```
data %>%
  mutate(Decision = fct_relevel(Decision, "Admit", "Waitlist", "Decline")) %>%
  group_by(Decision) %>%
  ggplot() + geom_boxplot(aes(Decision, GPA)) + theme_bw()
```
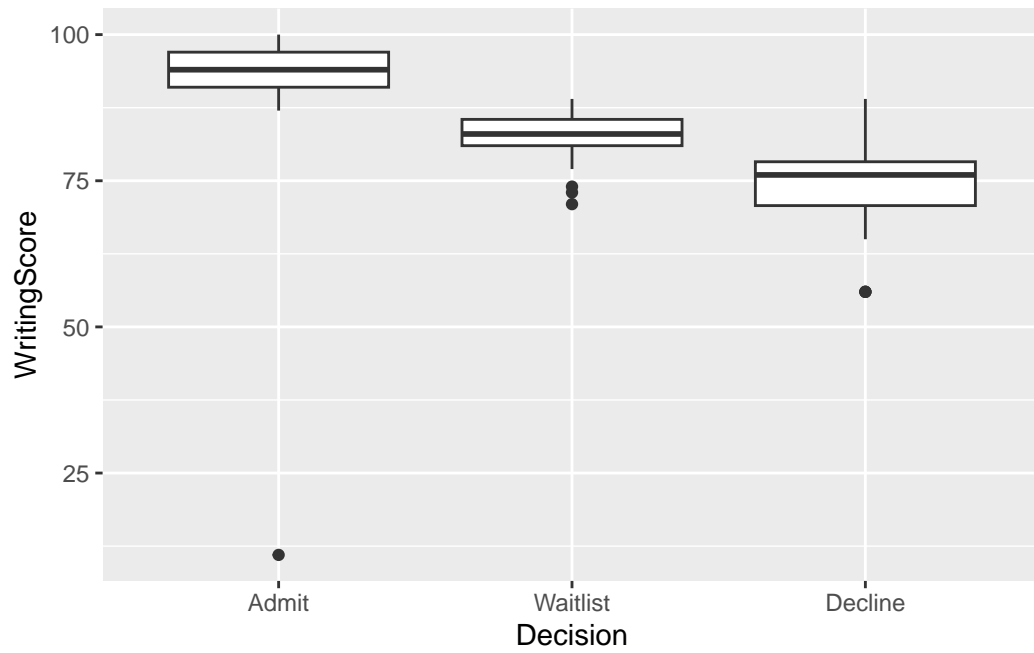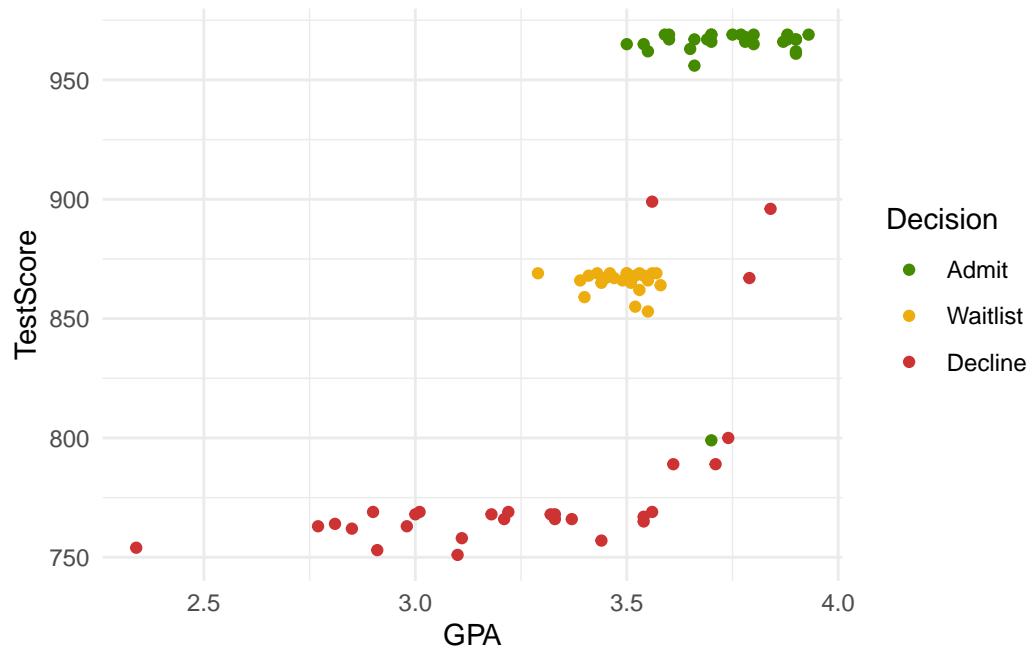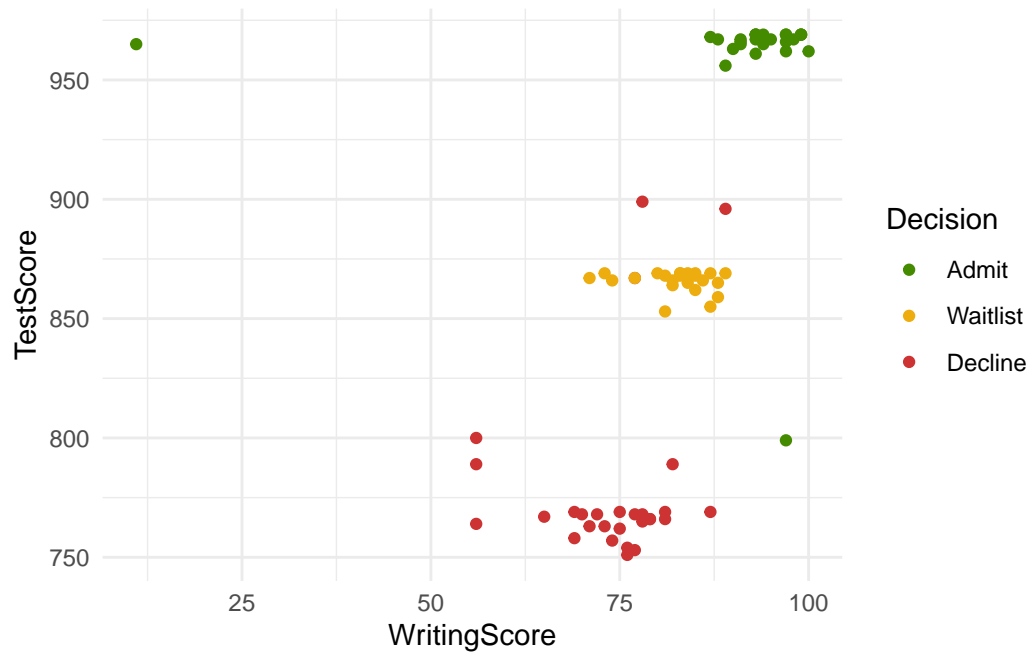
```
data %>%
  mutate(Decision = fct_relevel(Decision, "Admit", "Waitlist", "Decline")) %>%
  group_by(Decision) %>%
  ggplot() + geom_boxplot(aes(Decision, TestScore)) + theme_bw()
```

```
data %>%
  mutate(Decision = fct_relevel(Decision, "Admit", "Waitlist", "Decline")) %>%
  group_by(Decision) %>%
  ggplot() + geom_boxplot(aes(Decision, WritingScore))
```

```
data %>%
  mutate(Decision = fct_relevel(Decision, "Admit", "Waitlist", "Decline")) %>%
  ggplot() + geom_point(aes(x=GPA, y=TestScore, color = Decision)) + scale_color_manual(va
```

```
data %>%
  mutate(Decision = fct_relevel(Decision, "Admit", "Waitlist", "Decline")) %>%
  ggplot() + geom_point(aes(x=WritingScore, y=TestScore, color = Decision)) + scale_color_
```

```
data %>%
  mutate(Decision = fct_relevel(Decision, "Admit", "Waitlist", "Decline")) %>%
  ggplot() + geom_point(aes(x=GPA, y=WritingScore, color = Decision)) + scale_color_manual
```
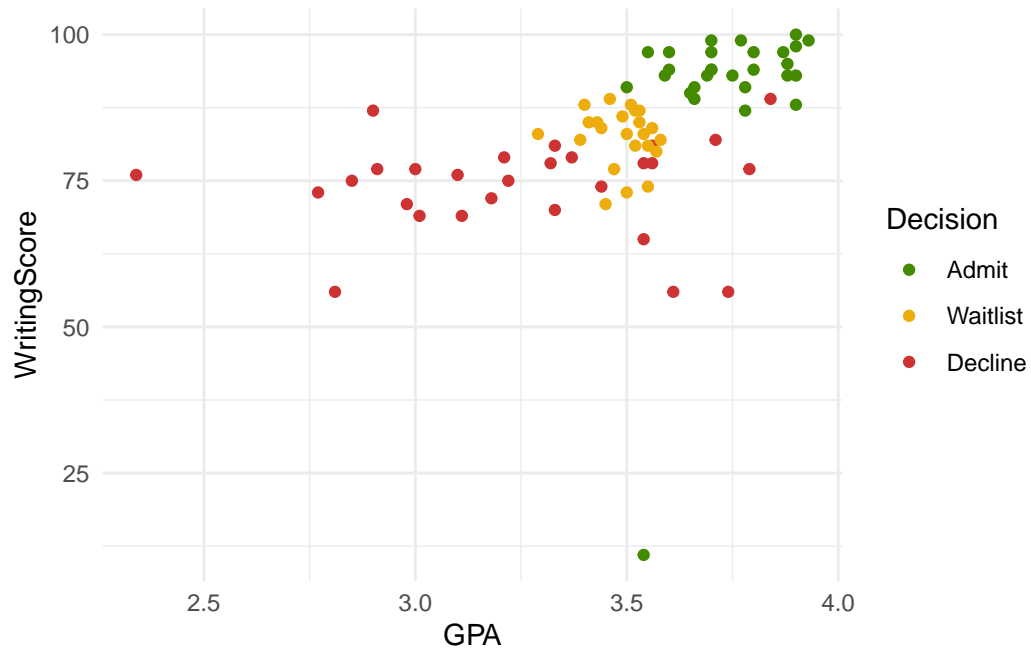
```r
library(corrr)
data %>%
  subset(select = -X) %>%
  correlate() %>%
  focus(TestScore) %>%
  gt() %>%
  tab_style(
    style = list(
      cell_fill(color = "seagreen2"),
      cell_text(weight = "bold")
      ),
    locations = cells_body(
      columns = TestScore,
      rows = TestScore >= 0.6
    )) %>%
  tab_style(
    style = list(
      cell_fill(color = "coral1"),
      cell_text(weight = "bold")
      ),
    locations = cells_body(
      columns = TestScore,
```

```
    rows = TestScore <= -0.1
  )) %>%
cols_label(
  term = "Variable",
  TestScore = "Correlation",
) %>%
tab_header(
  title = md("Correlation of Test Score with all other variables "),
)
```

```
Non-numeric variables removed from input: `Decision`, and `State`
Correlation computed with
* Method: 'pearson'
* Missing treated using: 'pairwise.complete.obs'
```

Correlation of Test Score with all other variables

Variable	Correlation
GPA	0.74183107
WorkExp	-0.03932081
WritingScore	0.52554018
Gender	0.04006486
VolunteerLevel	-0.13291980