423 final viz

```
library(readr)
```

```
Warning: package 'readr' was built under R version 4.2.3
```

```
library(ggplot2)
```

```
Warning: package 'ggplot2' was built under R version 4.2.3
```

```
library(tidyverse)
```

```
Warning: package 'dplyr' was built under R version 4.2.3

Warning: package 'stringr' was built under R version 4.2.3

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ stringr   1.5.1
✔ forcats   1.0.0     ✔ tibble    3.2.1
✔ lubridate 1.9.3     ✔ tidyr     1.3.0
✔ purrr     1.0.2
── 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
become errors
```

```
sleep <- read_csv("~/Downloads/sleep_cycle_productivity.csv")
```

```
Rows: 5000 Columns: 15
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr   (1): Gender
dbl  (13): Person_ID, Age, Sleep Start Time, Sleep End Time, Total Sleep Hou...
date  (1): Date

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
n_distinct(sleep$Person_ID)
```

```
[1] 3858
```

```
nrow(sleep)
```

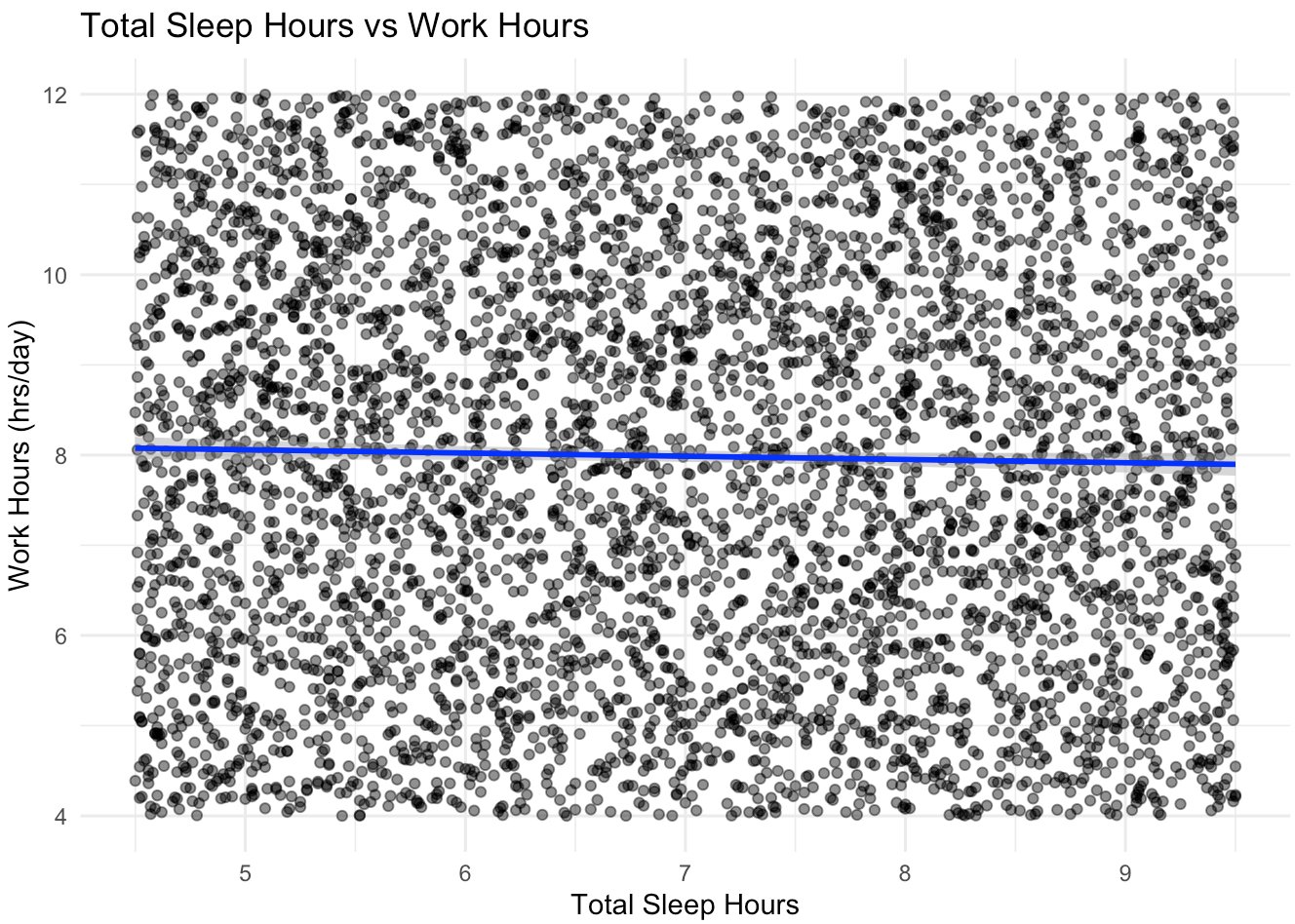
```
[1] 5000
```

```r
sleep %>% group_by(Person_ID)
```

```
# A tibble: 5,000 × 15
# Groups:   Person_ID [3,858]
   Date       Person_ID   Age Gender `Sleep Start Time` `Sleep End Time`
   <date>         <dbl> <dbl> <chr>                <dbl>            <dbl>
 1 2024-04-12      1860    32 Other                 23.3             4.61
 2 2024-11-04      1769    41 Female                21.0             2.43
 3 2024-08-31      2528    20 Male                  22.1             3.45
 4 2024-02-22      8041    37 Other                 23.1             6.65
 5 2024-02-23      4843    46 Other                 21.4             4.17
 6 2024-07-08      7439    38 Male                  21.8             6.41
 7 2024-01-09      6463    18 Other                 22.8             6.87
 8 2024-01-28      7278    26 Female                20.8             3.14
 9 2024-04-10      9110    31 Other                 20.1             3.37
10 2024-02-21      6116    49 Female                20.4             3.89
# ℹ 4,990 more rows
# ℹ 9 more variables: `Total Sleep Hours` <dbl>, `Sleep Quality` <dbl>,
#   `Exercise (mins/day)` <dbl>, `Caffeine Intake (mg)` <dbl>,
#   `Screen Time Before Bed (mins)` <dbl>, `Work Hours (hrs/day)` <dbl>,
#   `Productivity Score` <dbl>, `Mood Score` <dbl>, `Stress Level` <dbl>
```

```r
ggplot(sleep, aes(x = `Total Sleep Hours`, y = `Work Hours (hrs/day)`)) +
  geom_point(alpha = 0.5) +  # Scatter points with some transparency
  geom_smooth(method = "lm", col = "blue", se = TRUE) +  # Add a trend line
  labs(title = "Total Sleep Hours vs Work Hours",
       x = "Total Sleep Hours",
       y = "Work Hours (hrs/day)") +
  theme_minimal()
```

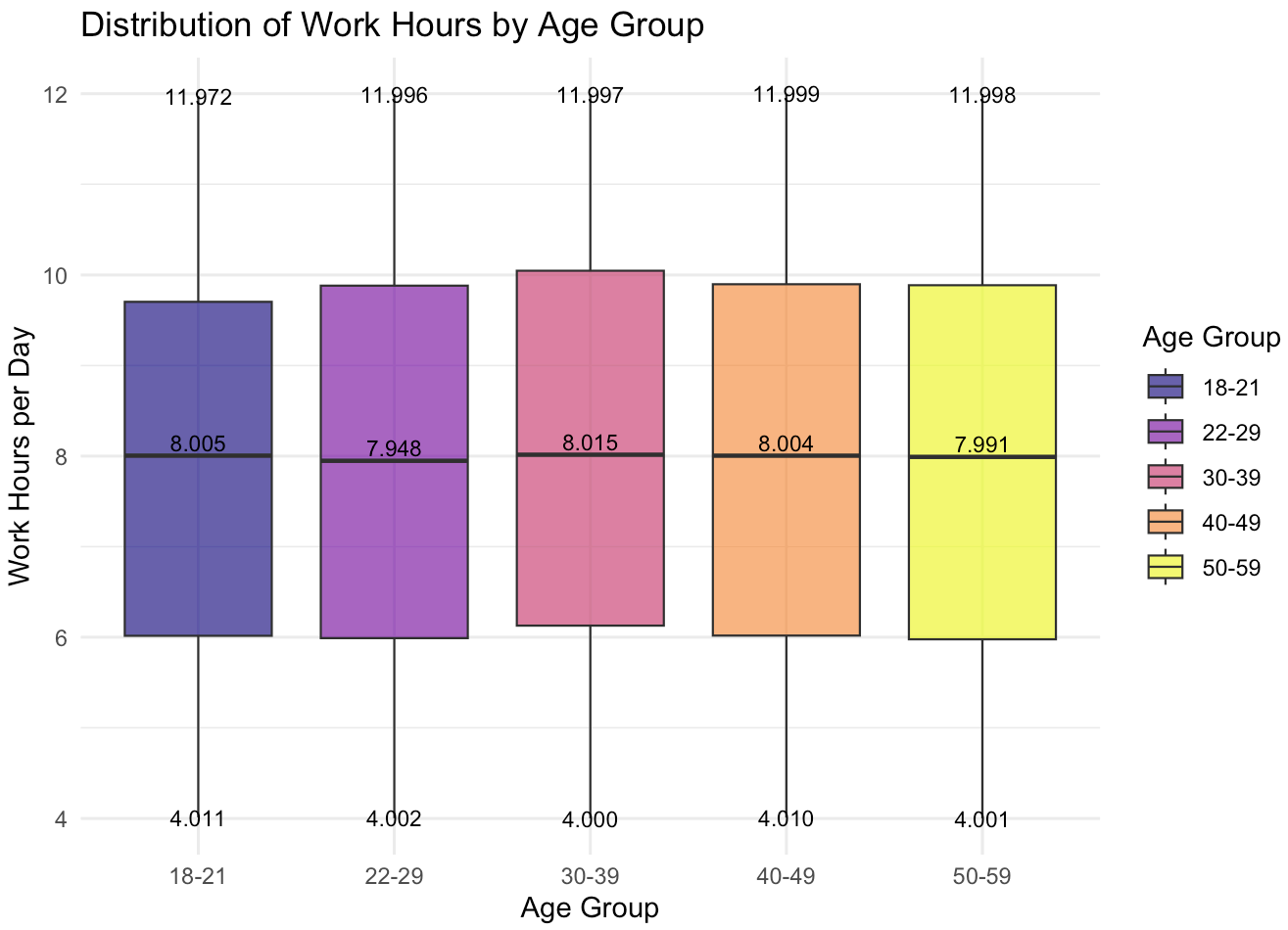
```
`geom_smooth()` using formula = 'y ~ x'
```




```r
# Create Age Categories
sleep <- sleep %>%
  mutate(Age_Group = cut(Age,
                         breaks = c(17, 21, 29, 39, 49, 59),
                         labels = c("18-21", "22-29", "30-39", "40-49", "50-59"))) %>%
  select(Date, Person_ID, Age, Age_Group, everything())

# Summarize data to get min, max, and mean Work Hours
# for each Sleep Hour within each Age Group
sleep_summary <- sleep %>%
  group_by(Age_Group) %>%
  summarise(
    Avg_Exercise_Hours = mean(`Exercise (mins/day)`),
    Min_Work_Hours = mean(min(`Work Hours (hrs/day)`, na.rm = TRUE)),
    Max_Work_Hours = max(`Work Hours (hrs/day)`, na.rm = TRUE),
    Mean_Work_Hours = mean(`Work Hours (hrs/day)`, na.rm = TRUE),
    .groups = "drop"
  )


# box plot of work hours by age group
ggplot(sleep, aes(x = Age_Group, y = `Work Hours (hrs/day)`, fill = Age_Group)) +
  geom_boxplot(alpha = 0.7, size = 0.4, outlier.size = 2, outlier.color = "black") +
```

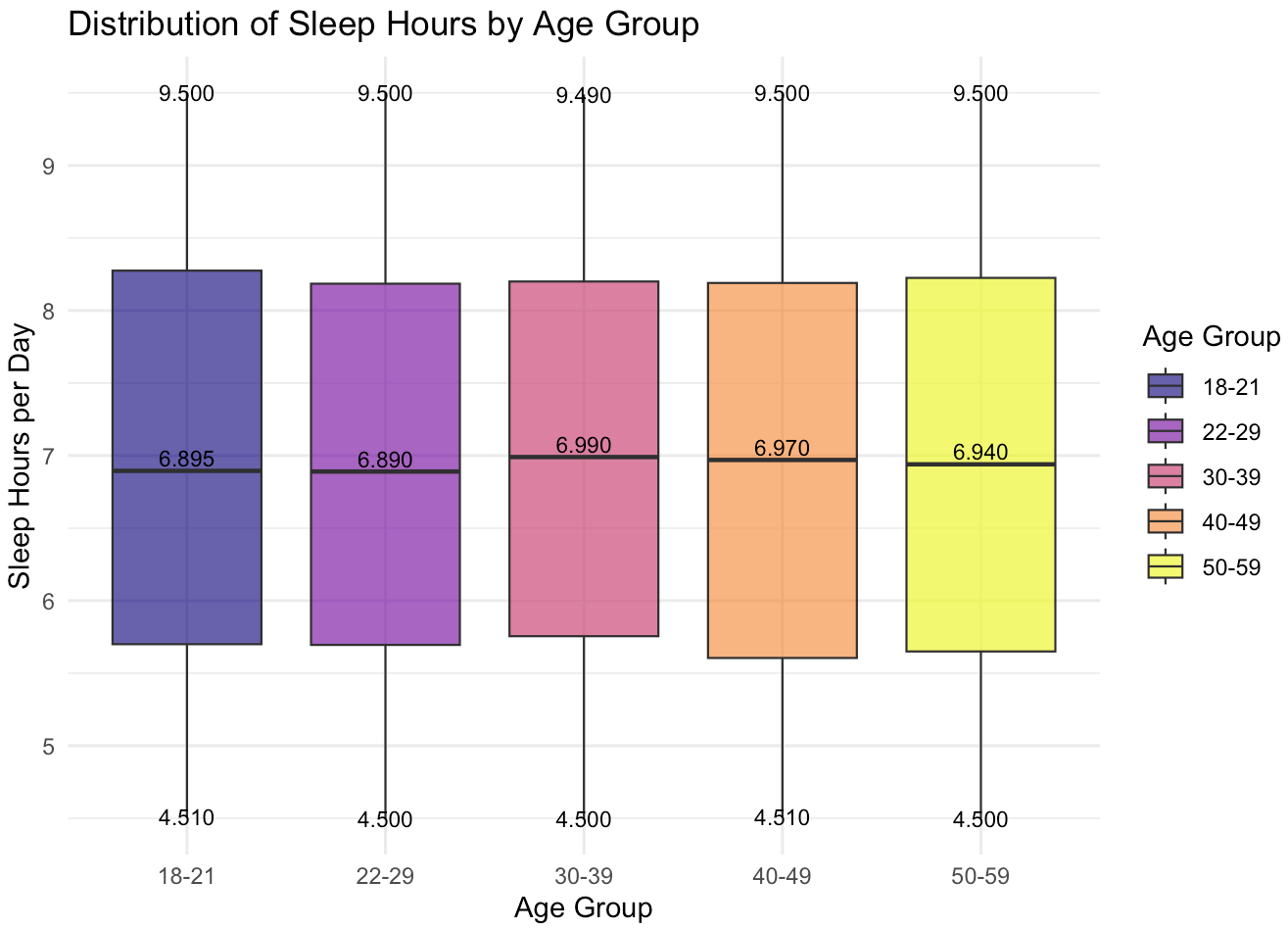
```
  stat_summary(fun = min, geom = "text", aes(label = sprintf("%.3f", ..y..)),
               nudge_y = -0.4, size = 3, color = "black") +
  stat_summary(fun = max, geom = "text", aes(label = sprintf("%.3f", ..y..)),
               nudge_y = 0.4, size = 3, color = "black") +
  stat_summary(fun = median, geom = "text", aes(label = sprintf("%.3f", ..y..)),
               vjust = -0.3, size = 3) +
  scale_fill_viridis_d(option = "plasma") +
  labs(title = "Distribution of Work Hours by Age Group",
       x = "Age Group",
       y = "Work Hours per Day",
       fill = "Age Group") +
  theme_minimal()
```

```
Warning in stat_summary(fun = min, geom = "text", aes(label = sprintf("%.3f", :
Ignoring unknown parameters: `nudge_y`

Warning in stat_summary(fun = max, geom = "text", aes(label = sprintf("%.3f", :
Ignoring unknown parameters: `nudge_y`

Warning: The dot-dot notation (`..y..`) was deprecated in ggplot2 3.4.0.
ℹ Please use `after_stat(y)` instead.
```

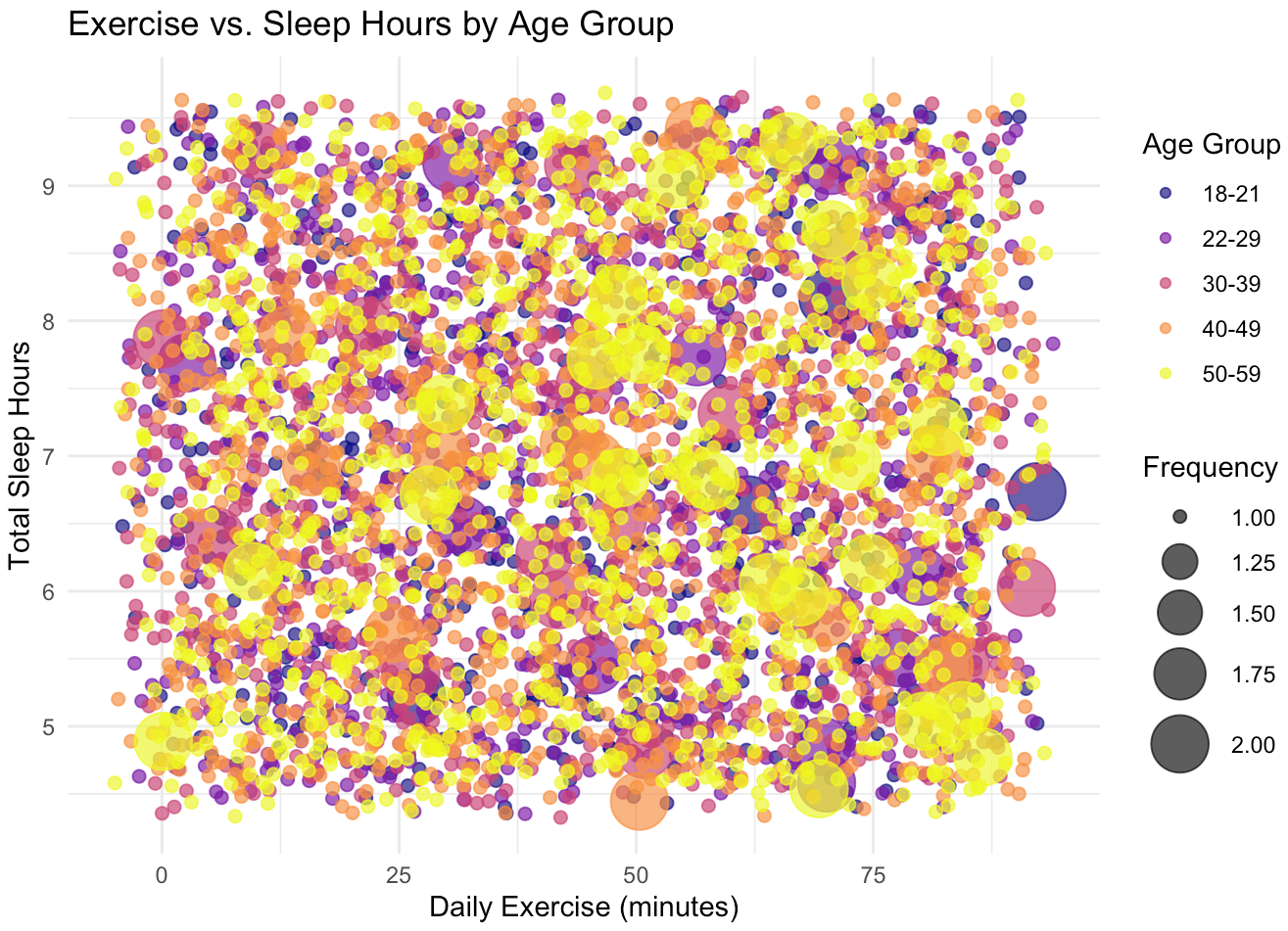


```r
# box plot of sleep hours by age group
ggplot(sleep, aes(x = Age_Group, y = `Total Sleep Hours`, fill = Age_Group)) +
  geom_boxplot(alpha = 0.7, size = 0.4, outlier.size = 2, outlier.color = "black") +
  stat_summary(fun = min, geom = "text", aes(label = sprintf("%.3f", ..y..)),
               nudge_y = -0.4, size = 3, color = "black") +
  stat_summary(fun = max, geom = "text", aes(label = sprintf("%.3f", ..y..)),
               nudge_y = 0.4, size = 3, color = "black") +
  stat_summary(fun = median, geom = "text", aes(label = sprintf("%.3f", ..y..)),
               vjust = -0.3, size = 3) +
  scale_fill_viridis_d(option = "plasma") +
  labs(title = "Distribution of Sleep Hours by Age Group",
       x = "Age Group",
       y = "Sleep Hours per Day",
       fill = "Age Group") +
  theme_minimal()
```

```
Warning in stat_summary(fun = min, geom = "text", aes(label = sprintf("%.3f", :
Ignoring unknown parameters: `nudge_y`

Warning in stat_summary(fun = max, geom = "text", aes(label = sprintf("%.3f", :
Ignoring unknown parameters: `nudge_y`
```



```r
# Load required libraries
library(ggplot2)
library(dplyr)

# Aggregate data to count occurrences (bubble size)
sleep_bubble <- sleep %>%
  group_by(Age_Group, `Exercise (mins/day)`, `Total Sleep Hours`) %>%
  summarise(Count = n(), .groups = "drop")  # Count occurrences for bubble size

# Create the bubble plot
ggplot(sleep_bubble, aes(x = `Exercise (mins/day)`,
                         y = `Total Sleep Hours`,
                         size = Count,
                         color = Age_Group)) +
  geom_point(alpha = 0.7, position = position_jitter(width = 5, height = 0.2)) +  # Bubbl
  scale_size(range = c(2, 10)) +  # Adjust bubble sizes
  scale_color_viridis_d(option = "plasma") +  # Color gradient for Age Group
  labs(title = "Exercise vs. Sleep Hours by Age Group",
       x = "Daily Exercise (minutes)",
       y = "Total Sleep Hours",
       size = "Frequency",
       color = "Age Group") +
  theme_minimal()
```



```r
sleep_summary <- sleep %>%
  group_by(Age_Group) %>%
  summarise(Average_Work_Hours = mean(`Work Hours (hrs/day)`, na.rm = TRUE),
            Average_Sleep_Hours = mean(`Total Sleep Hours`, na.rm = TRUE)) %>%
  pivot_longer(cols = c(Average_Work_Hours, Average_Sleep_Hours),
               names_to = "Category", values_to = "Hours")

ggplot(sleep_summary, aes(x = Age_Group, y = Hours, fill = Category)) +
  geom_bar(stat = "identity", position = "dodge", alpha = 0.8) +
  geom_text(aes(label = sprintf("%.3f", Hours)),
            position = position_dodge(width = 0.9), vjust = -0.5, size = 3) +
  scale_fill_manual(values = c("Average_Work_Hours" = "#FF9999", "Average_Sleep_Hours" =
  scale_y_continuous(limits = c(0, 9), breaks = seq(0, 9, by = 1))  +
  labs(title = "Average Work Hours and Sleep Hours by Age Group",
       x = "Age Group",
       y = "Average Hours",
       fill = "Category") +
  theme_minimal()
```