

# Survey

Garcia, Piano, Talaban, Ticot

2025-12-15

## Load packages

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.6
## v forcats    1.0.1      v stringr   1.5.2
## v ggplot2    4.0.1      v tibble    3.3.0
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.1.0
## -- 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 become errors
```

```
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test
```

```
library(ggplot2)
```

```
library(dplyr)
```

```
theme_set(
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
)
```

```
#Import
```

```
df_raw <- read_csv("dataset.csv")
```

```
## Rows: 55 Columns: 10
## -- Column specification -----
## Delimiter: ","
## chr (4): Timestamp, Year Level, Average number of hours spent studying per d...
## dbl (6): Age, Number of subjects you are currently enrolled in, How many tim...
##
## 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.
```

```

df_raw <- as.data.frame(df_raw)

#Clean
df <- df_raw |>
  clean_names() |>
  subset(select = -c(timestamp, feedback_about_our_survey))

#Make sure df is a data.frame
df <- as.data.frame(df)

#Type conversions
df <- df |>
  mutate(
    gadgets_hours_week = as.numeric(on_average_how_many_hours_per_week_do_you_spend_using_gadgets_for_s
    age = as.numeric(age),
    number_of_subjects_you_are_currently_enrolled_in = as.numeric(number_of_subjects_you_are_currently_
    how_many_times_per_week_do_you_review_your_lessons = as.numeric(how_many_times_per_week_do_you_revi
    how_many_times_per_month_do_you_submit_homework_late = as.numeric(how_many_times_per_month_do_you_s
    how_many_clubs_or_organizations_are_you_involved_in = as.numeric(how_many_clubs_or_organizations_ar
    year_level = as.factor(year_level)
  )

# Ensure final object is data.frame
df <- as.data.frame(df)

#Descriptive statistics
statistic <- df |>
  summarise(
    n = n(),
    mean_age = mean(age, na.rm = TRUE),
    median_age = median(age, na.rm = TRUE),
    min_age = min(age, na.rm = TRUE),
    max_age = max(age, na.rm = TRUE),

    mean_subjects = mean(number_of_subjects_you_are_currently_enrolled_in, na.rm = TRUE),
    median_subjects = median(number_of_subjects_you_are_currently_enrolled_in, na.rm = TRUE),
    min_subjects = min(number_of_subjects_you_are_currently_enrolled_in, na.rm = TRUE),
    max_subjects = max(number_of_subjects_you_are_currently_enrolled_in, na.rm = TRUE),

    mean_review = mean(how_many_times_per_week_do_you_review_your_lessons, na.rm = TRUE),
    median_review = median(how_many_times_per_week_do_you_review_your_lessons, na.rm = TRUE),
    min_review = min(how_many_times_per_week_do_you_review_your_lessons, na.rm = TRUE),
    max_review = max(how_many_times_per_week_do_you_review_your_lessons, na.rm = TRUE),

    mean_late = mean(how_many_times_per_month_do_you_submit_homework_late, na.rm = TRUE),
    median_late = median(how_many_times_per_month_do_you_submit_homework_late, na.rm = TRUE),
    min_late = min(how_many_times_per_month_do_you_submit_homework_late, na.rm = TRUE),
    max_late = max(how_many_times_per_month_do_you_submit_homework_late, na.rm = TRUE),

    mean_clubs = mean(how_many_clubs_or_organizations_are_you_involved_in, na.rm = TRUE),
    median_clubs = median(how_many_clubs_or_organizations_are_you_involved_in, na.rm = TRUE),
    min_clubs = min(how_many_clubs_or_organizations_are_you_involved_in, na.rm = TRUE),
    max_clubs = max(how_many_clubs_or_organizations_are_you_involved_in, na.rm = TRUE),

    mean_gadgets = mean(gadgets_hours_week, na.rm = TRUE),

```

```

median_gadgets = median(gadgets_hours_week, na.rm = TRUE),
min_gadgets    = min(gadgets_hours_week, na.rm = TRUE),
max_gadgets    = max(gadgets_hours_week, na.rm = TRUE)
)

```

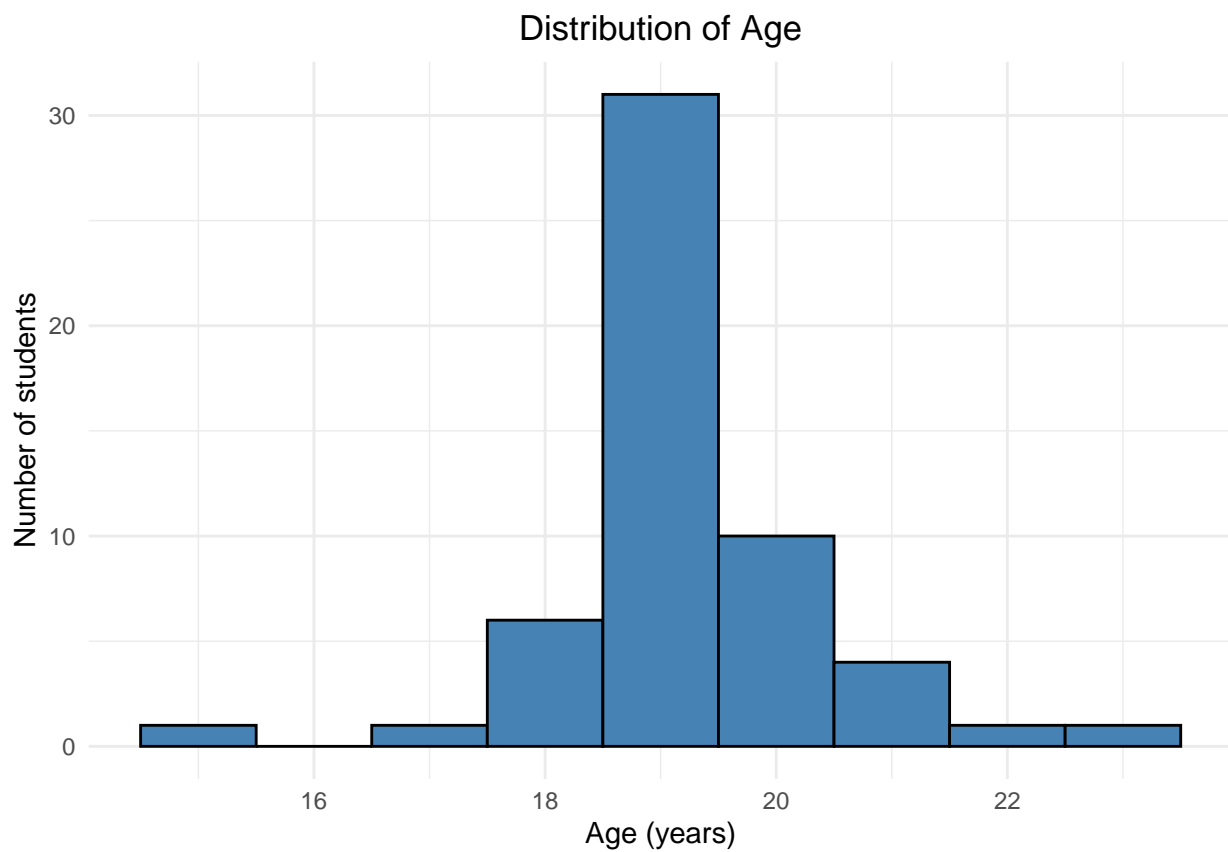
*#Plots*

*#Age*

```

ggplot(df, aes(x = age)) +
  geom_histogram(binwidth = 1, fill = "steelblue", color = "black") +
  labs(
    title = "Distribution of Age",
    x = "Age (years)",
    y = "Number of students"
  )
)

```

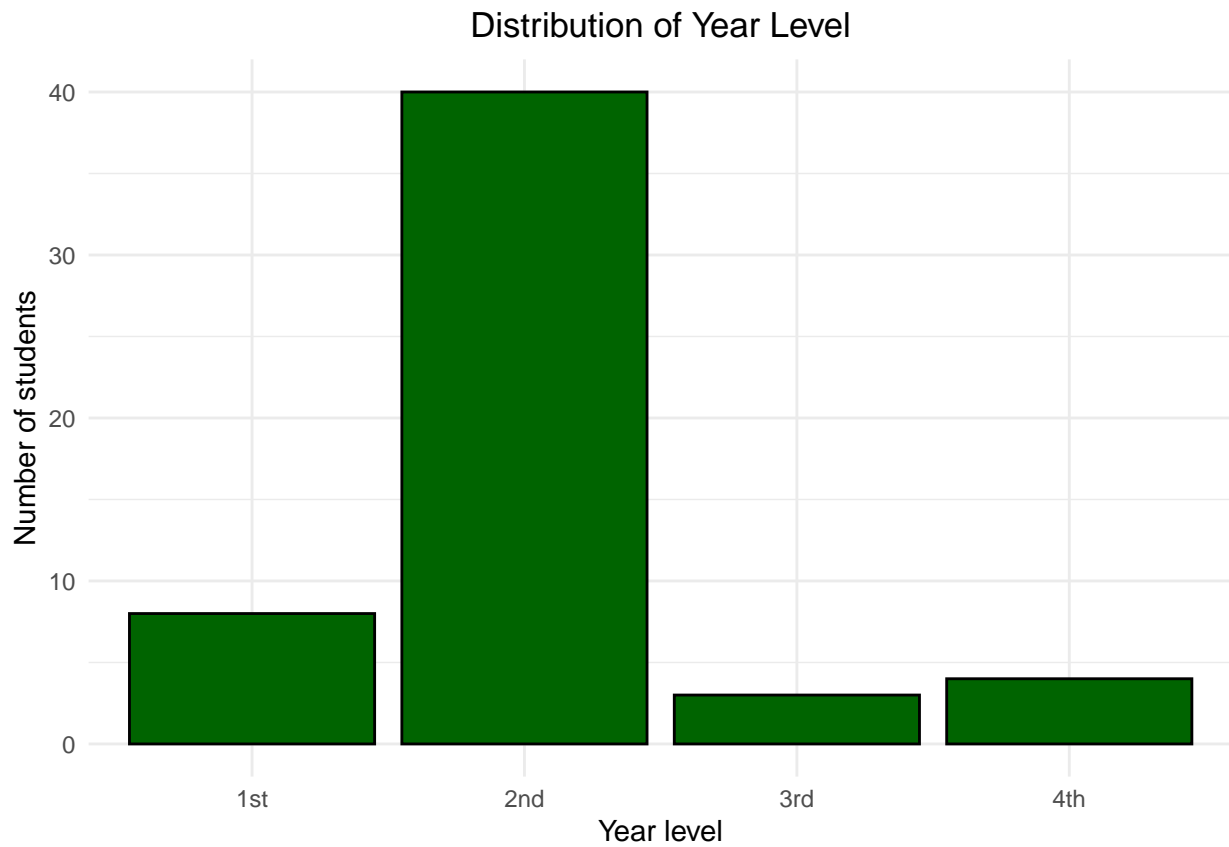


*#Year level*

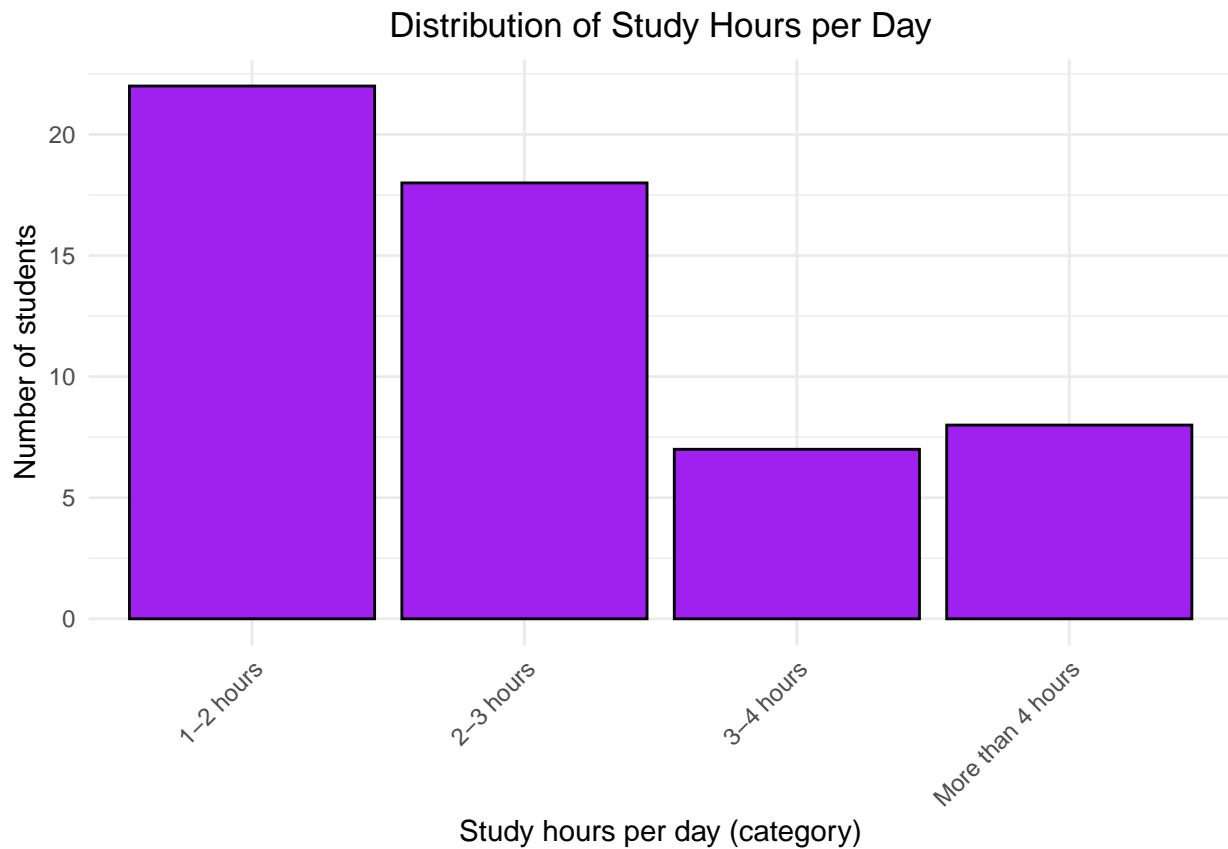
```

ggplot(df, aes(x = year_level)) +
  geom_bar(fill = "darkgreen", color = "black") +
  labs(
    title = "Distribution of Year Level",
    x = "Year level",
    y = "Number of students"
  )
)

```

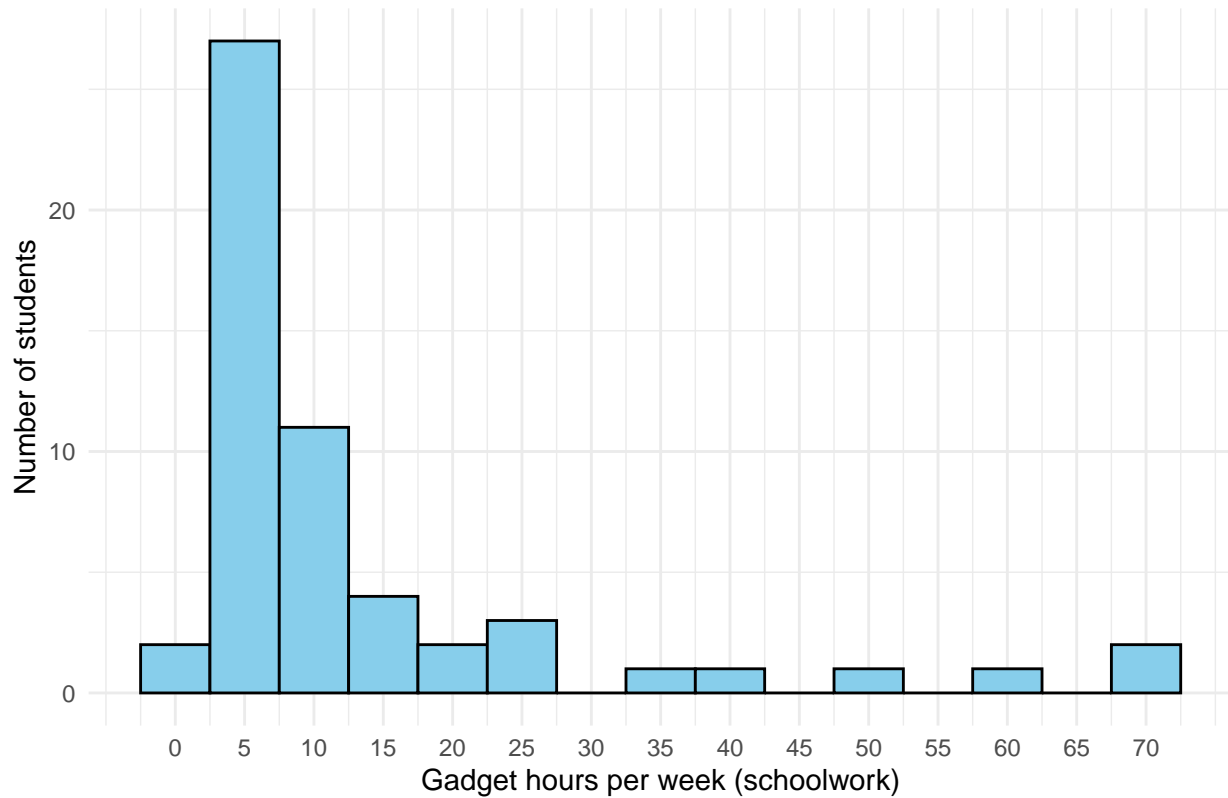


```
#Study hours per day (categorical)
ggplot(df, aes(x = average_number_of_hours_spent_studying_per_day)) +
  geom_bar(fill = "purple", color = "black") +
  labs(
    title = "Distribution of Study Hours per Day",
    x = "Study hours per day (category)",
    y = "Number of students"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

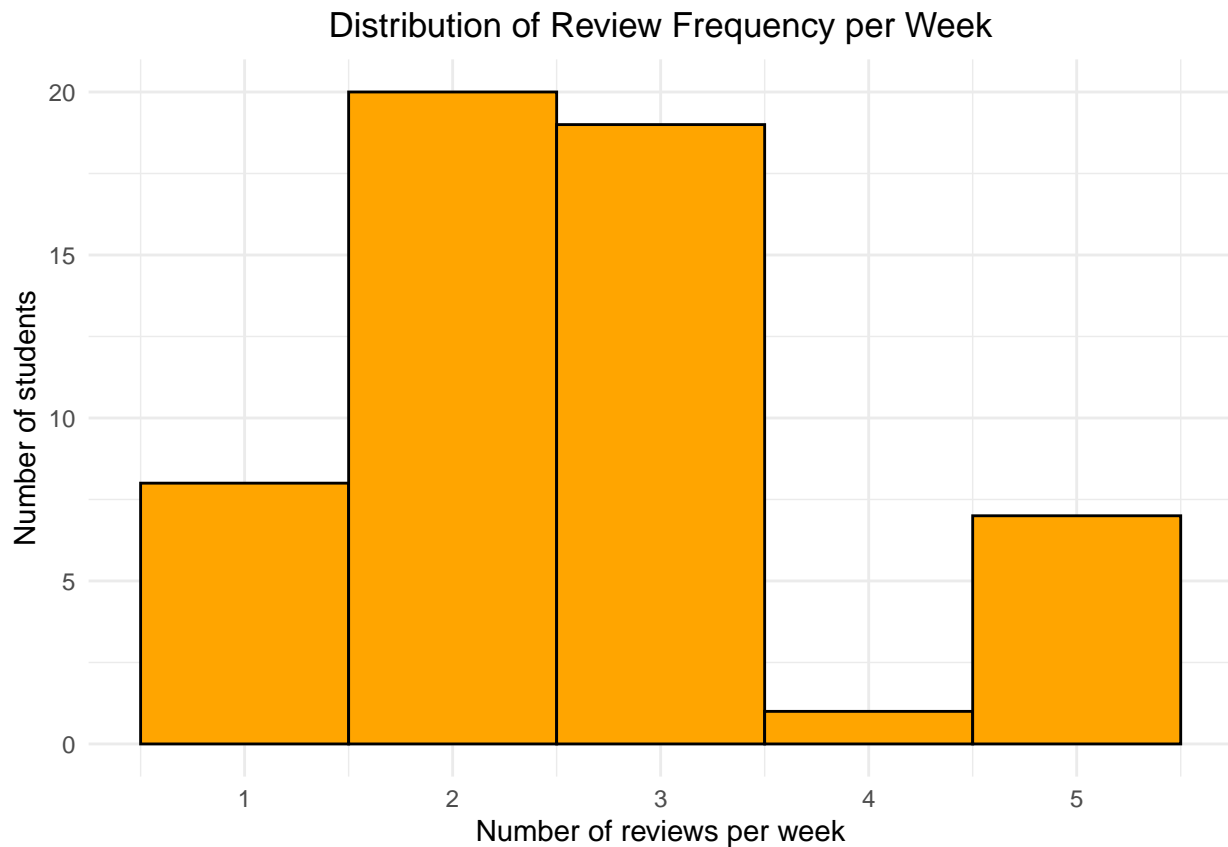


```
#Gadget hours per week
ggplot(df, aes(x = gadgets_hours_week)) +
  geom_histogram(binwidth = 5, fill = "skyblue", color = "black") +
  labs(
    title = "Distribution of Gadget Hours per Week for Schoolwork",
    x = "Gadget hours per week (schoolwork)",
    y = "Number of students"
  ) +
  scale_x_continuous(
    breaks = seq(
      0,
      max(df$gadgets_hours_week, na.rm = TRUE),
      by = 5
    )
  )
)
```

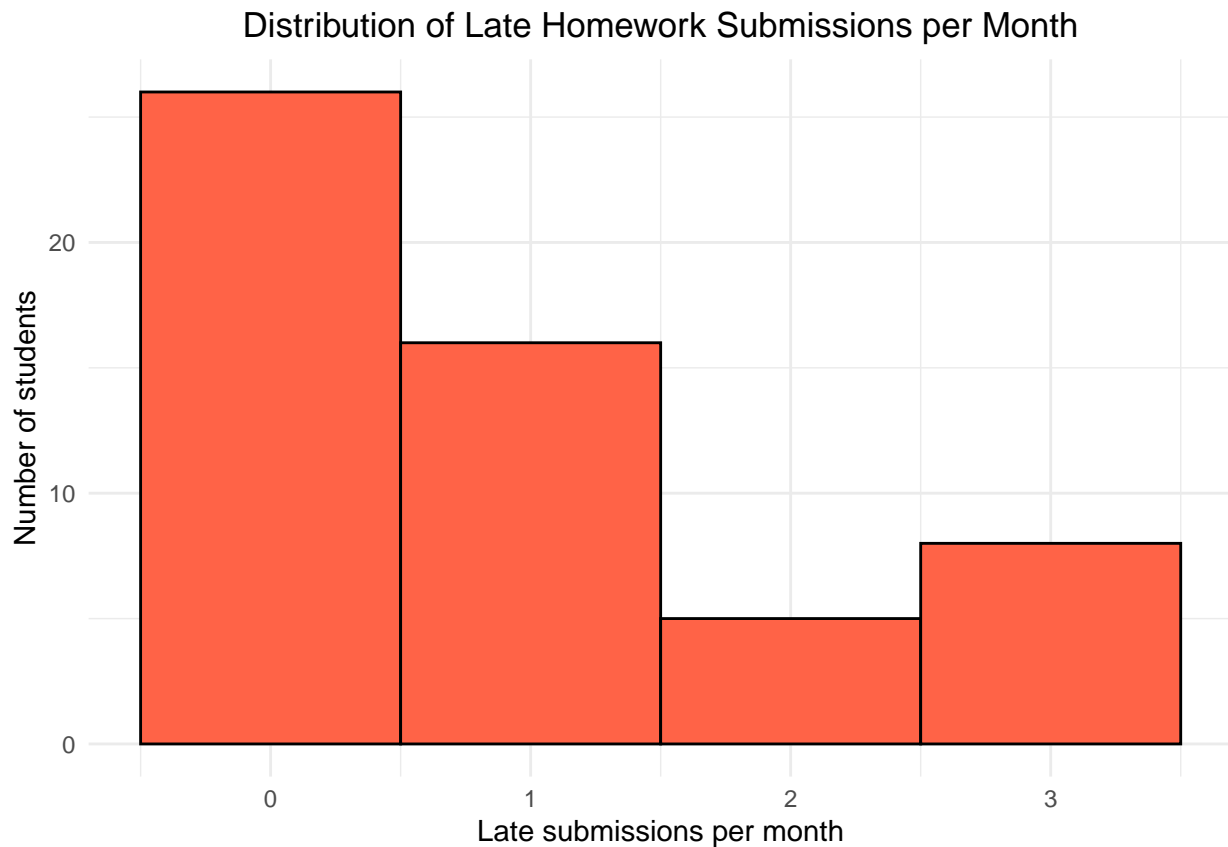
Distribution of Gadget Hours per Week for Schoolwork



```
#Review frequency per week
ggplot(df, aes(x = how_many_times_per_week_do_you_review_your_lessons)) +
  geom_histogram(binwidth = 1, fill = "orange", color = "black") +
  labs(
    title = "Distribution of Review Frequency per Week",
    x = "Number of reviews per week",
    y = "Number of students"
  ) +
  scale_x_continuous(
    breaks = seq(
      0,
      max(df$how_many_times_per_week_do_you_review_your_lessons, na.rm = TRUE),
      by = 1
    )
  )
)
```

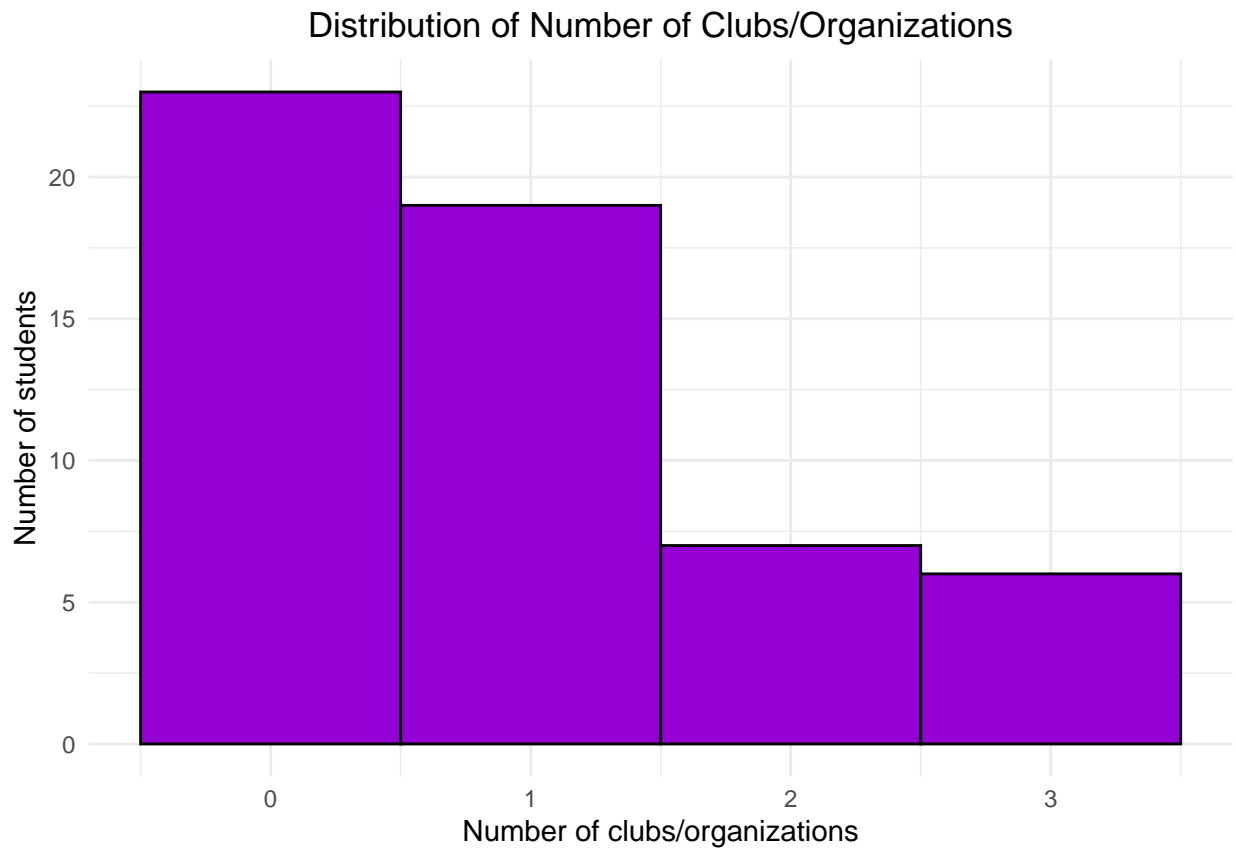


```
#Late submissions per month
ggplot(df, aes(x = how_many_times_per_month_do_you_submit_homework_late)) +
  geom_histogram(binwidth = 1, fill = "tomato", color = "black") +
  labs(
    title = "Distribution of Late Homework Submissions per Month",
    x = "Late submissions per month",
    y = "Number of students"
  ) +
  scale_x_continuous(
    breaks = seq(
      0,
      max(df$how_many_times_per_month_do_you_submit_homework_late, na.rm = TRUE),
      by = 1
    )
  )
)
```

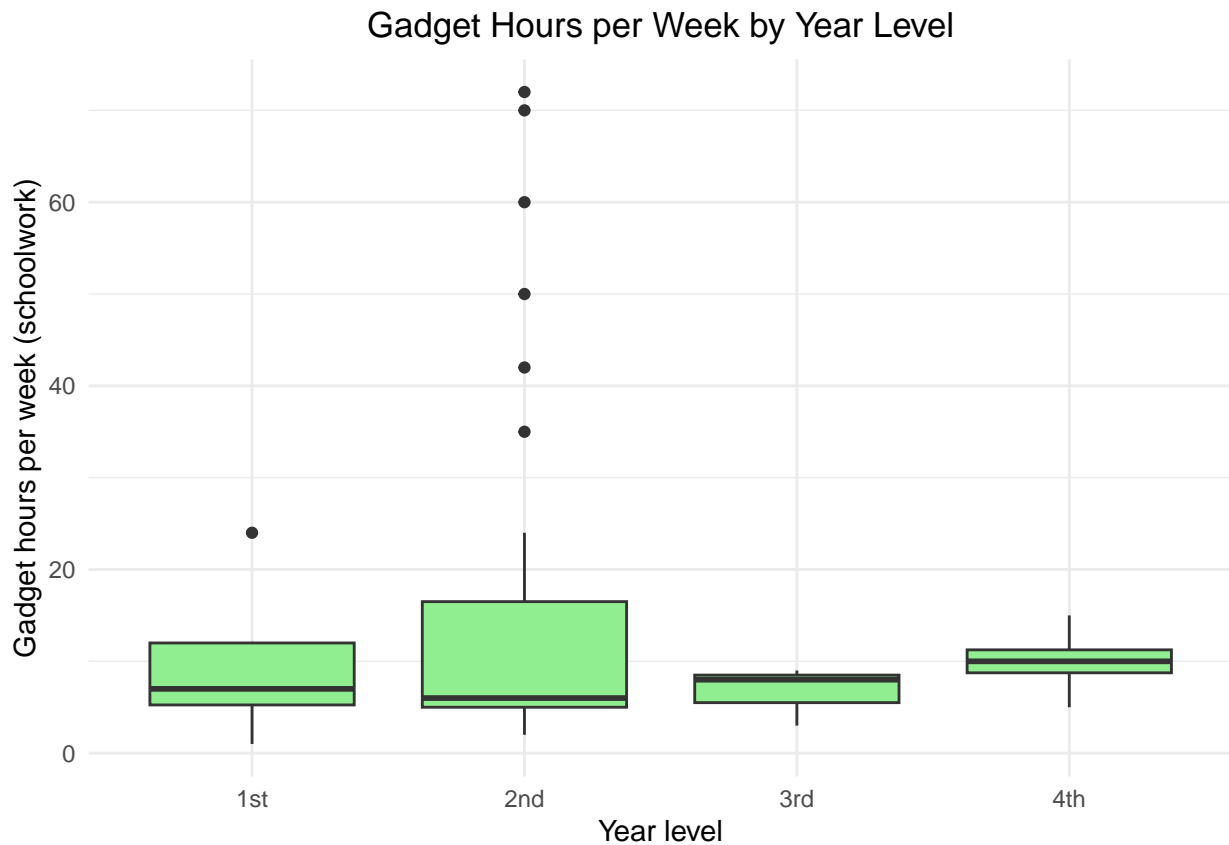


```
#Number of clubs
ggplot(df, aes(x = how_many_clubs_or_organizations_are_you_involved_in)) +
  geom_histogram(binwidth = 1, fill = "darkviolet", color = "black") +
  labs(
    title = "Distribution of Number of Clubs/Organizations",
    x = "Number of clubs/organizations",
    y = "Number of students"
  ) +
  scale_x_continuous(
    breaks = seq(
      0,
      max(df$how_many_clubs_or_organizations_are_you_involved_in, na.rm = TRUE),
      by = 1
    )
  )
```





```
#Gadget hours by year level  
ggplot(df, aes(x = year_level, y = gadgets_hours_week)) +  
  geom_boxplot(fill = "lightgreen") +  
  labs(  
    title = "Gadget Hours per Week by Year Level",  
    x = "Year level",  
    y = "Gadget hours per week (schoolwork)"  
  )
```



```
#Gadget hours by study-hours category
ggplot(df, aes(
  x = average_number_of_hours_spent_studying_per_day,
  y = gadgets_hours_week
)) +
  geom_boxplot(fill = "skyblue") +
  labs(
    title = "Gadget Hours per Week by Study Hours Category",
    x = "Study hours per day (category)",
    y = "Gadget hours per week (schoolwork)"
  ) +
  coord_flip()
```

