The suggested structure for the report is:

1. **Background/context of the business (100 words):** Briefly describe the business scenario (context) and the business problem you aim to solve for Turtle Games.
2. **Analytical approach (350 words):** Describe the approach taken to import, clean, and analyse data in Python and R. Include a detailed and insightful description of the processes you used and the decisions you made during analysis, such as the choice of libraries, functions, and variables. Ensure that the description of the steps taken to prepare data for analysis is clear, well-organised, and relevant to Turtle Games’s objectives.

The marketing department also wants to better understand the usefulness of remuneration and spending scores in providing data for analysis but does not know where to begin. You are tasked to identify groups within the customer base that can be used to target specific market segments.

1. **Visualisation and insights (350 words):** Describe the rationale for the selected visualisations, and ensure that the interpretations of the visualisation outputs are detailed, insightful, and relevant to the business objectives.
2. **Patterns and predictions (200 words):**Clearly articulate the patterns you’ve discovered and the predictions you’ve made. Describe how they relate to the business scenario, and address Turtle Games’s objectives.
3. **Ensure** that you use the correct naming convention before you submit. Save and submit your files as:
   * LastName\_FirstName\_DA301\_Assignment\_Report.pdf

A graph of a number of numbers

Description automatically generated with medium confidence

A graph showing a number of numbers

Description automatically generated with medium confidence

A graph showing a number of numbers

Description automatically generated

A graph showing a number of dots

Description automatically generated

A graph showing a graph of a sales rate

Description automatically generated with medium confidence

A graph showing a number of black dots

Description automatically generated A graph showing the value of a number of points

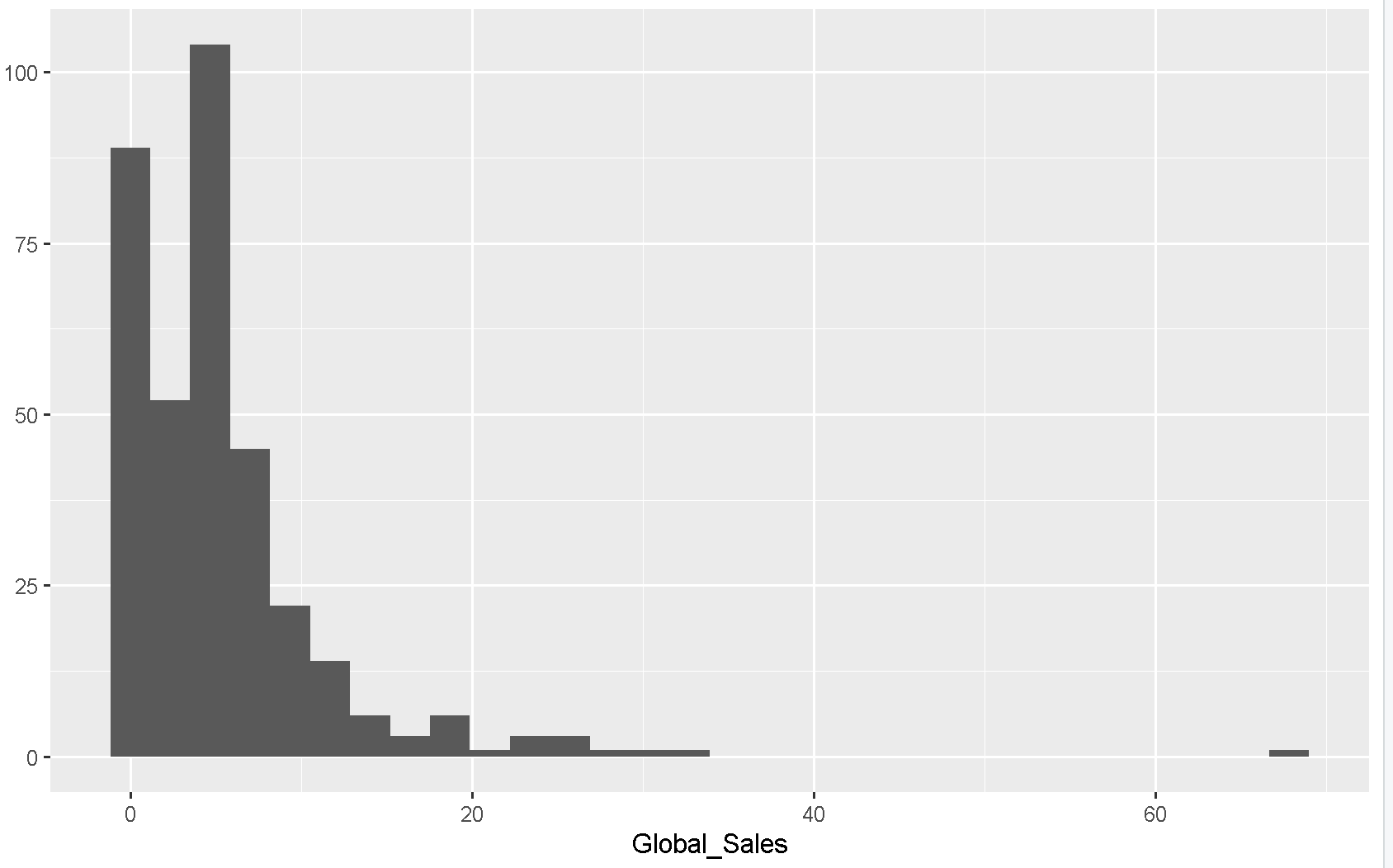
Description automatically generated with medium confidence

A graph showing a number of dots

Description automatically generated

A graph showing a number of dots

Description automatically generated



A graph of a graph

Description automatically generated

A graph of a sales graph

Description automatically generated with medium confidence

A graph with a number of lines

Description automatically generated with medium confidence A graph with a bar chart

Description automatically generated

A graph with a number of points

Description automatically generated with medium confidence

A graph with a graph of a graph

Description automatically generated with medium confidence

## 2b) Determine which plot is the best to compare game sales.

# Create scatterplots.

qplot(Platform,sum\_EU, data=df\_s\_p)

qplot(Platform,sum\_NA, data=df\_s)

qplot(Platform,sum\_global, data=df\_s)

qplot(y=sum\_global, data=df\_s\_p)

qplot(y=sum\_NA, data=df\_s\_p)

qplot(y=sum\_EU, data=df\_s\_p)

# Create histograms.

qplot(sum\_global, bins=5, data=df\_s\_p)

qplot(sum\_global, data=df\_s\_p)

qplot(sum\_EU, data=df\_s\_p)

qplot(sum\_NA, data=df\_s\_p)

# Create boxplots.

qplot(sum\_global, Platform, data=df\_s\_p, geom='boxplot')

qplot(sum\_NA, Platform, data=df\_s\_p, geom='boxplot')

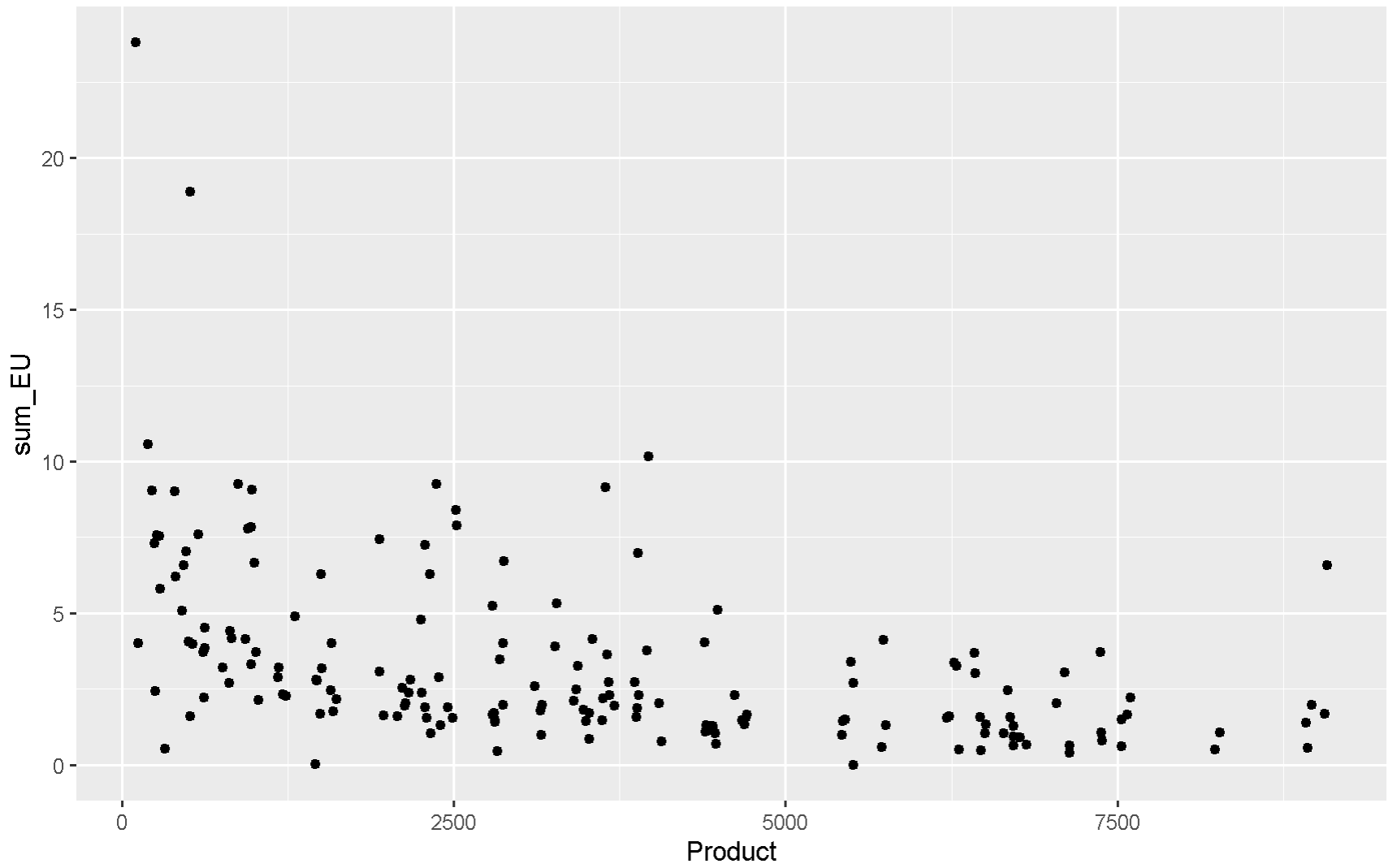
qplot(sum\_EU, Platform, data=df\_s\_p, geom='boxplot')

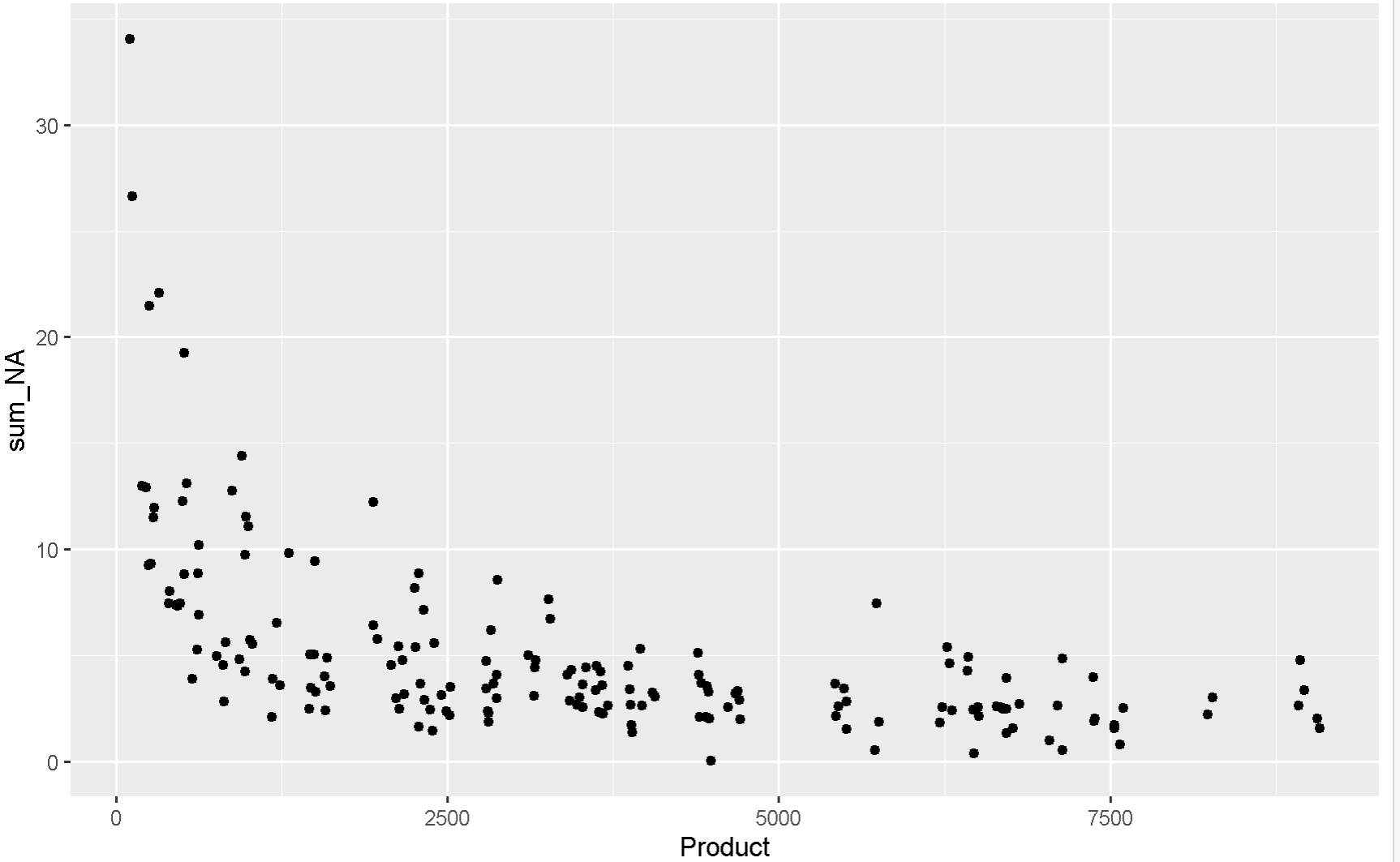
# Specify the qqnorm function.  
# Draw a qqplot using the total\_seconds data.  
qqnorm(drive3$total\_seconds,  
       col='blue',  
       xlab="z Value",  
       ylab='Time')  
  
# Specify the qqline function.  
# Add a reference line to the qqplot.  
qqline(drive3$total\_seconds,  
       col='red',  
       lwd=2)   
  
# Run a Shapiro-Wilk test.  
shapiro.test(drive3$total\_seconds)  
  
# Specify the skewness and kurtosis functions.  
skewness(drive3$total\_seconds)   
kurtosis(drive3$total\_seconds)

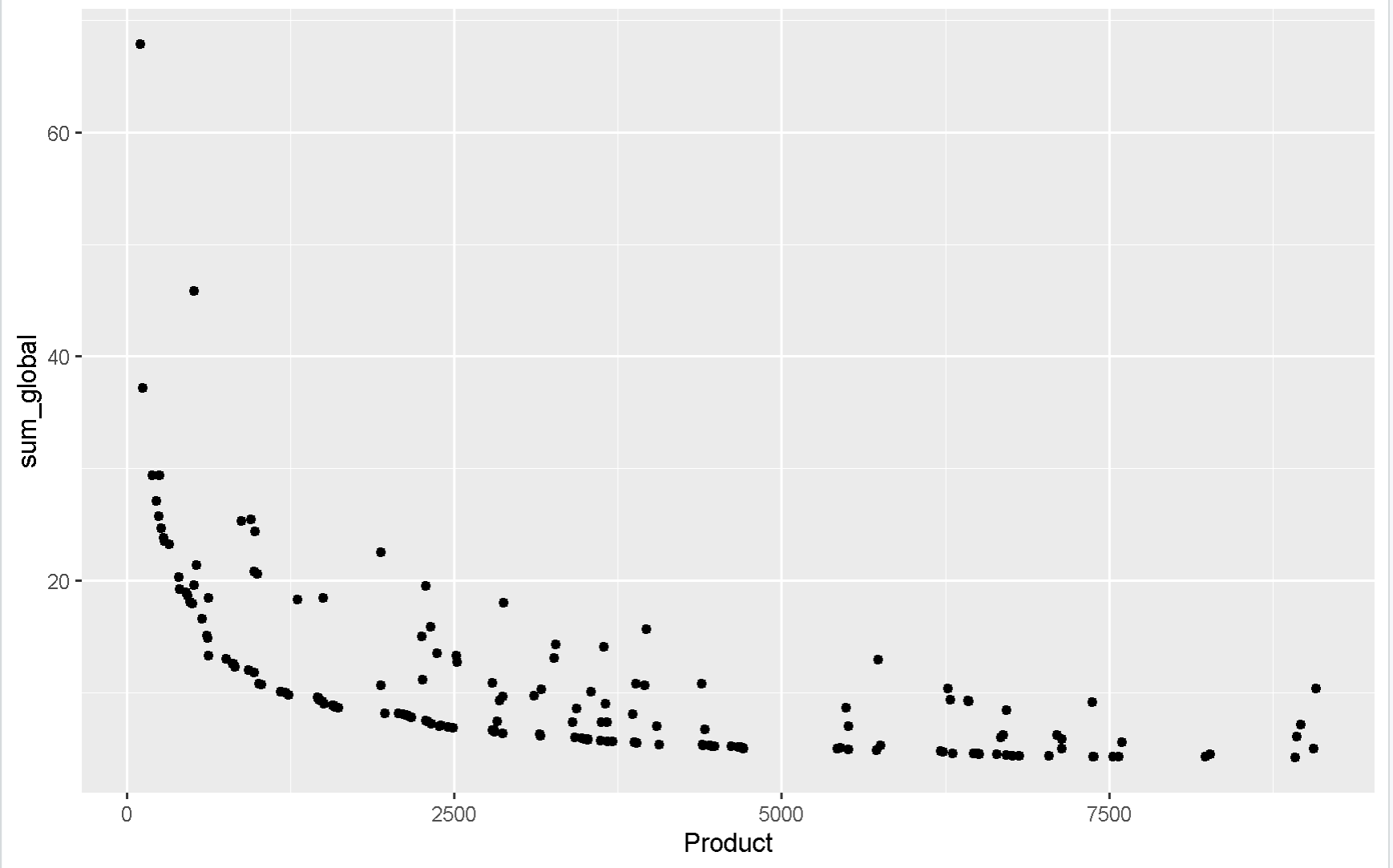
# Specify the t.test function.  
# Set the data source, the confidence interval (95%),  
# and the theoretical mean.   
  
t.test(drive3$total\_seconds,  
       conf.level=0.95,  
       mu=120)

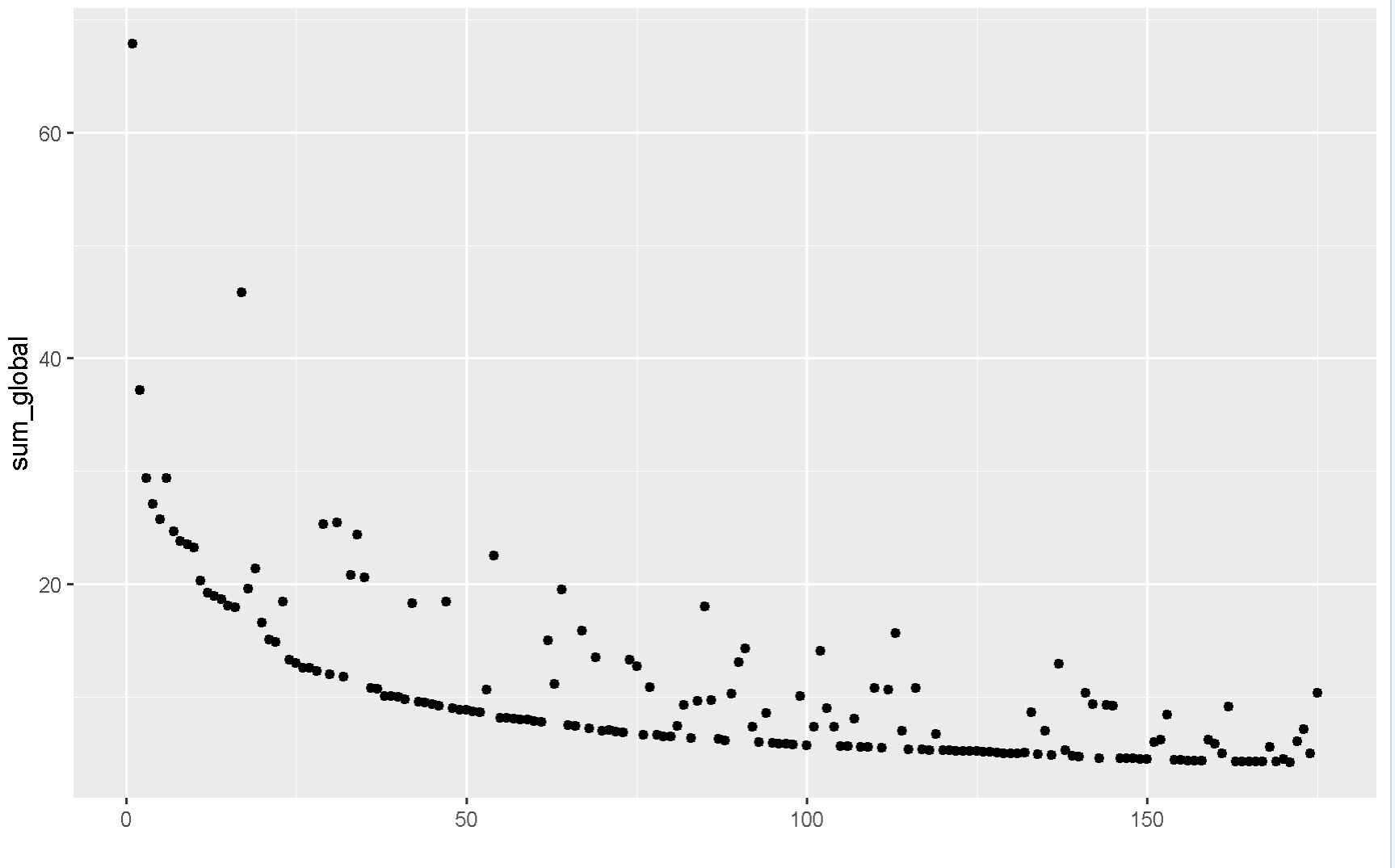
# Run a Shapiro-Wilk test.  
shapiro.test(drive3$car\_stop)  
shapiro.test(drive3$car\_go)  
shapiro.test(drive3$take\_order)  
shapiro.test(drive3$hand\_over\_order)

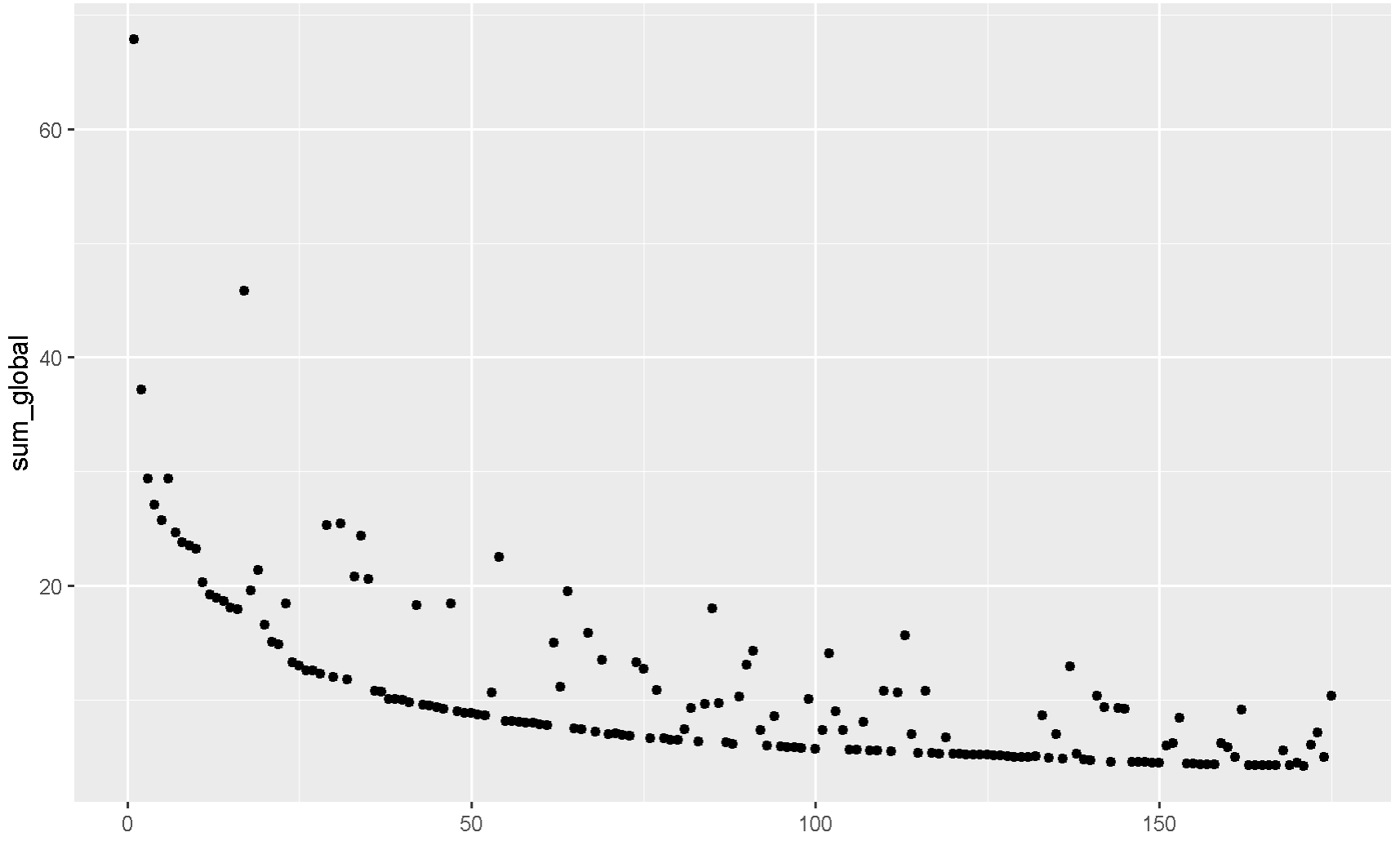
# Specify the cor function.  
# Set the first and second variables.  
  
cor(drive3$car\_stop, drive3$car\_go)  
cor(drive3$take\_order, drive3$hand\_over\_order)   
  
# Determine the correlation for the whole data frame.  
round (cor(drive3),  
       digits=2)

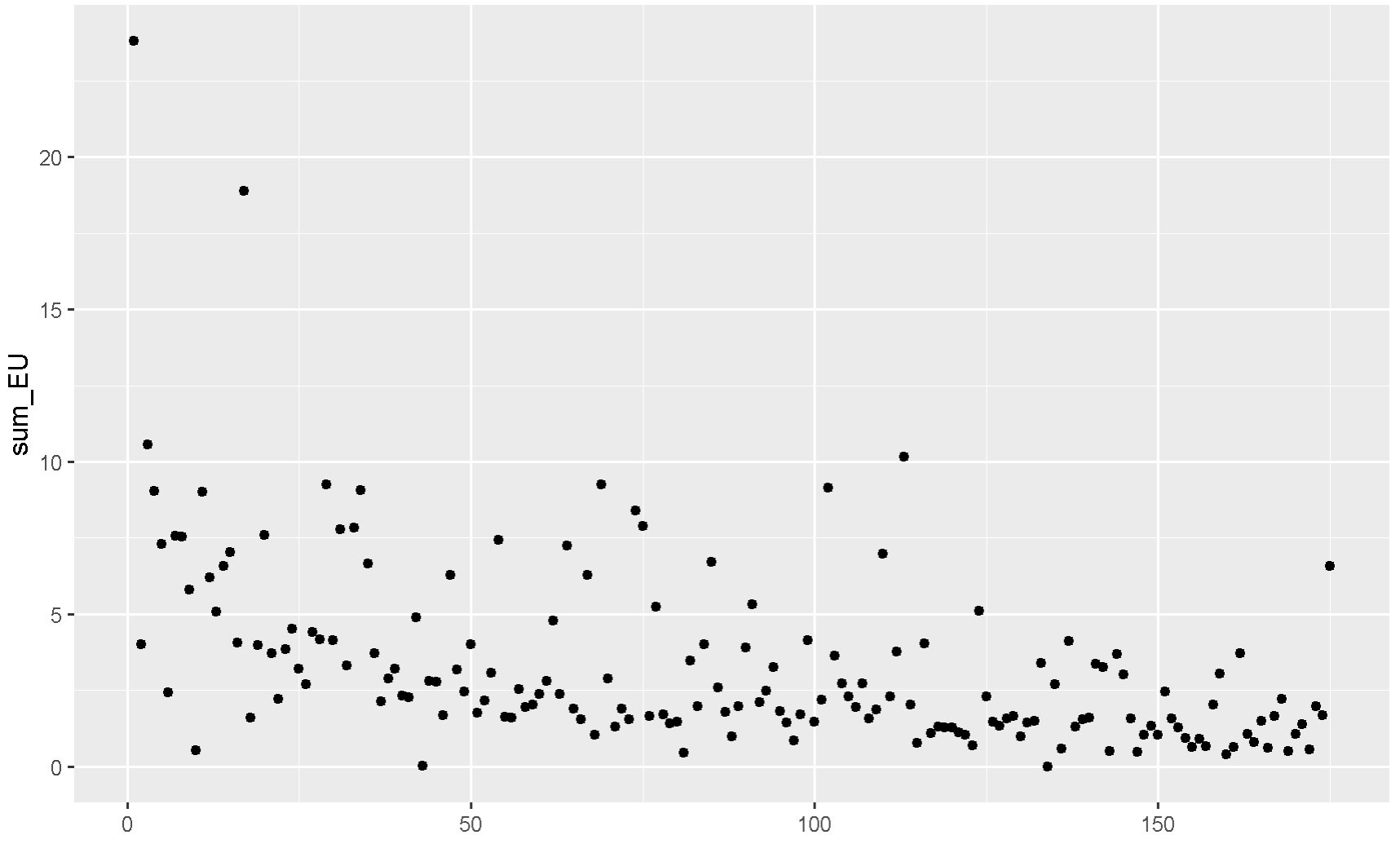


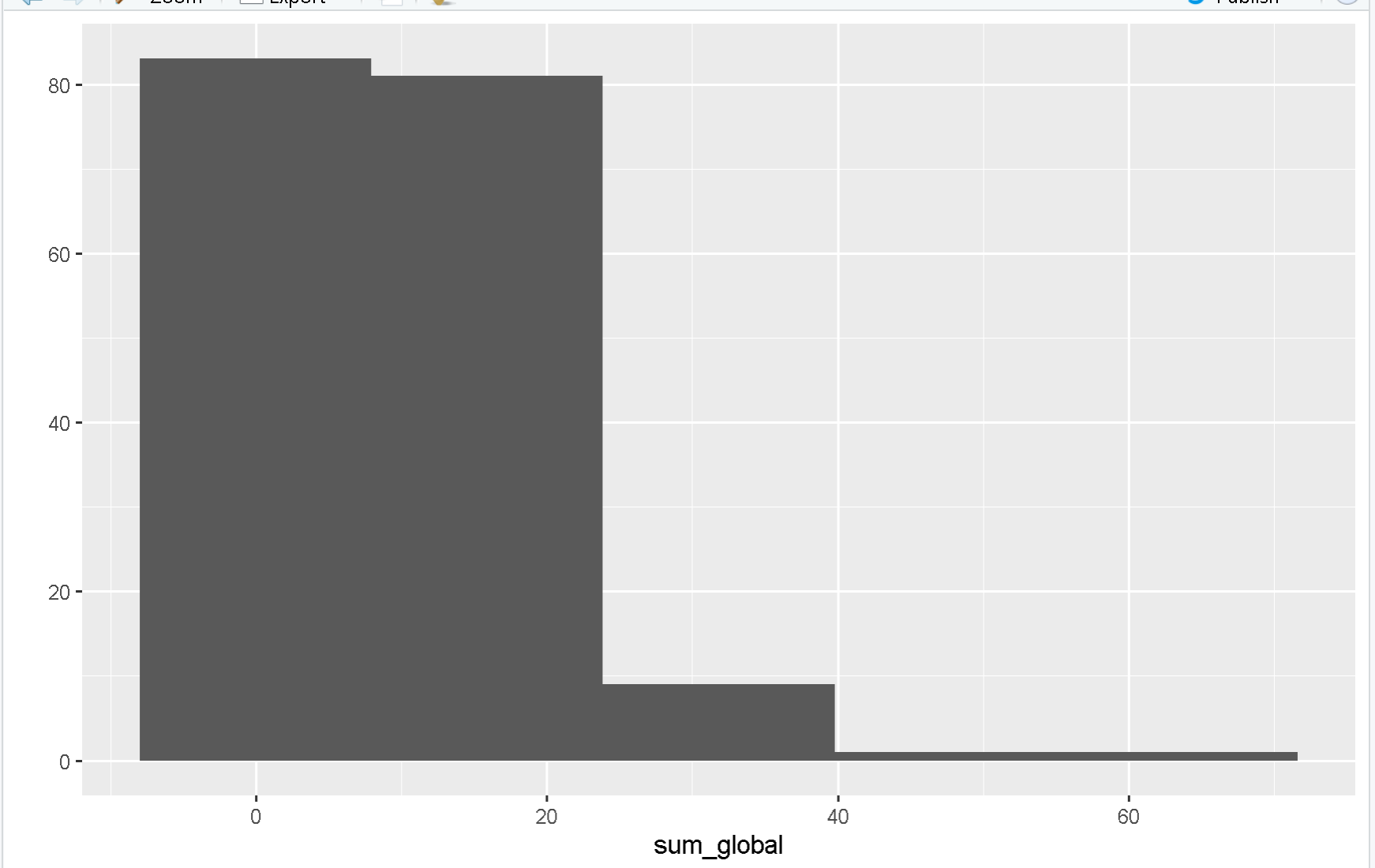


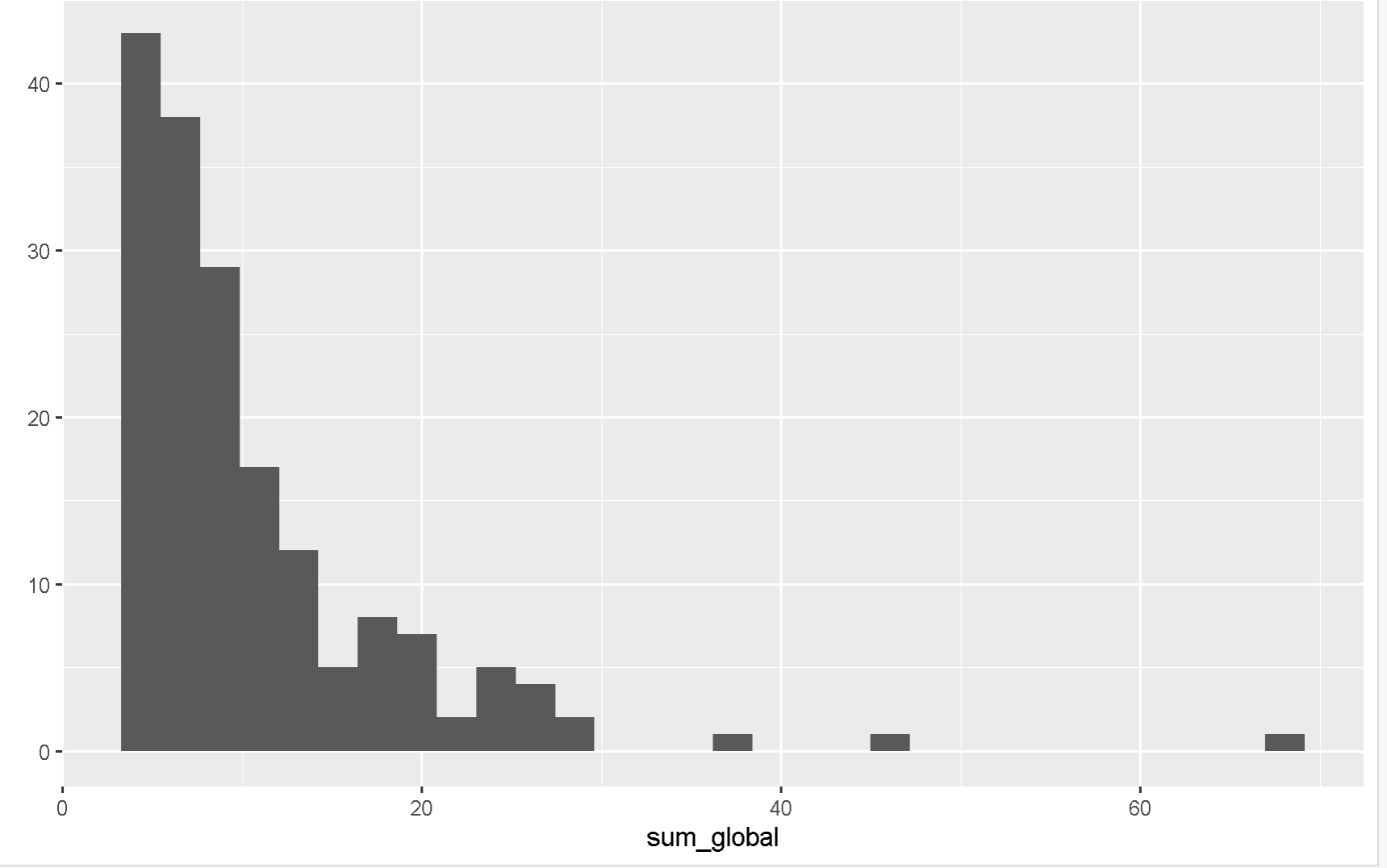


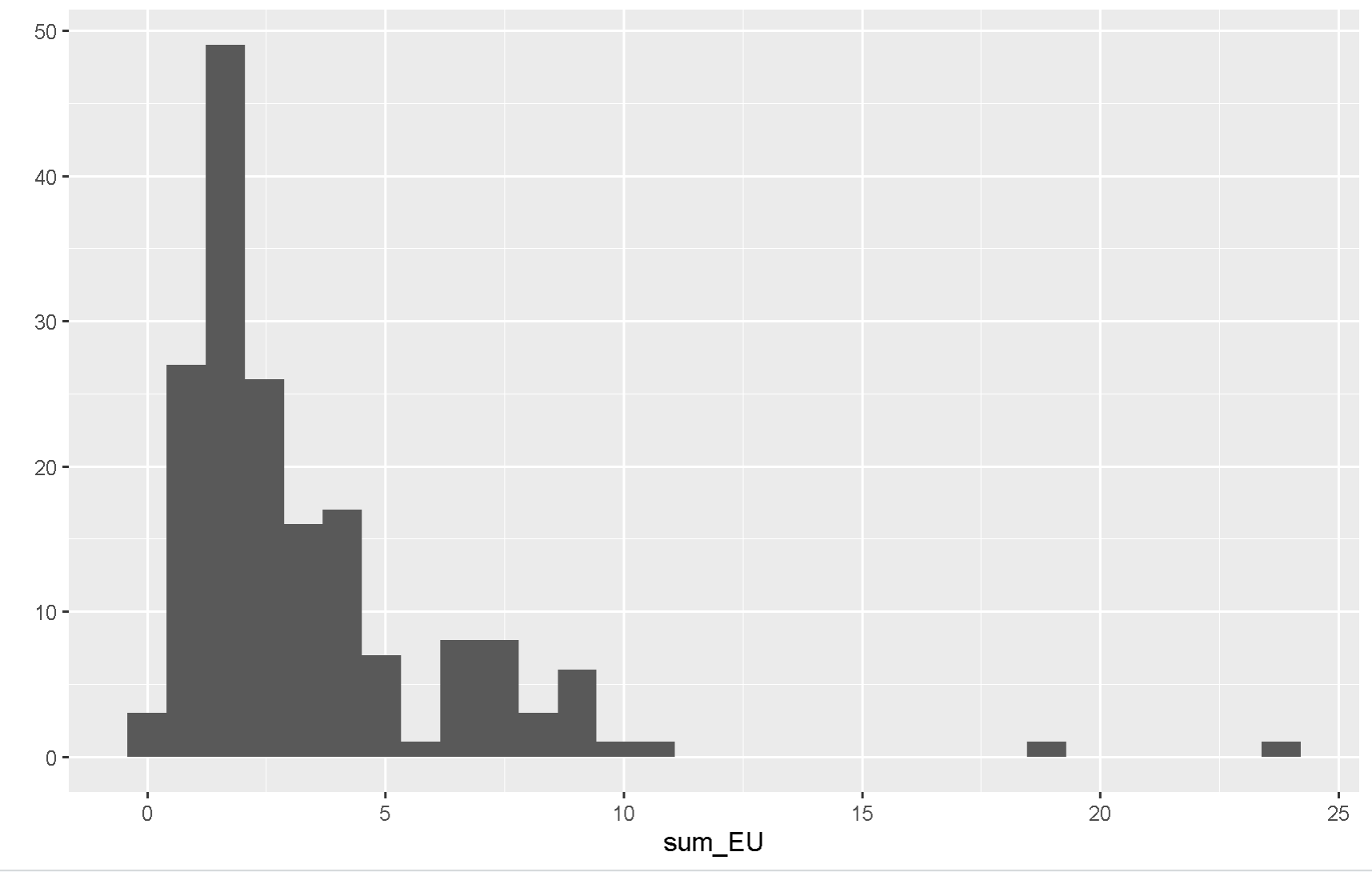


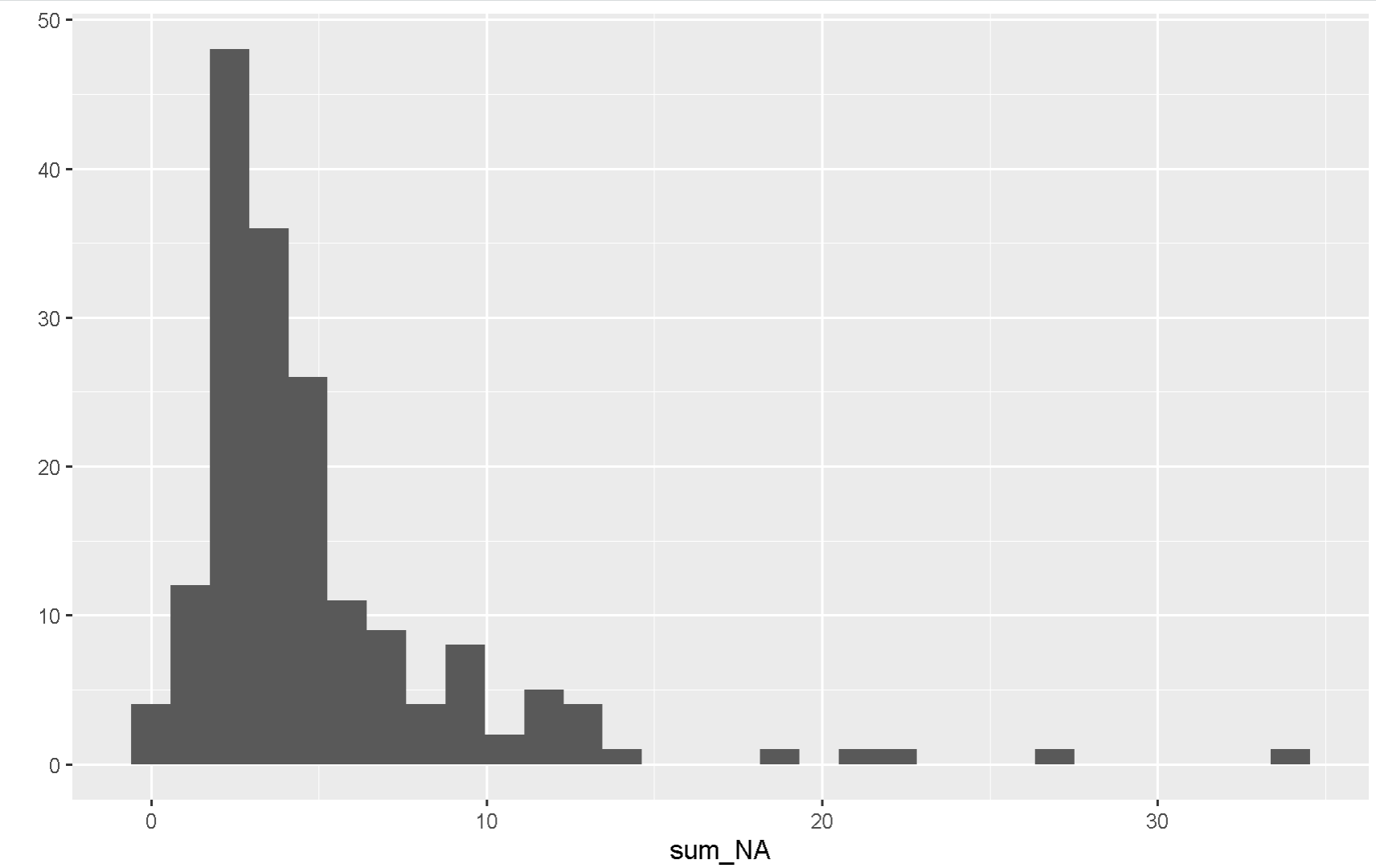


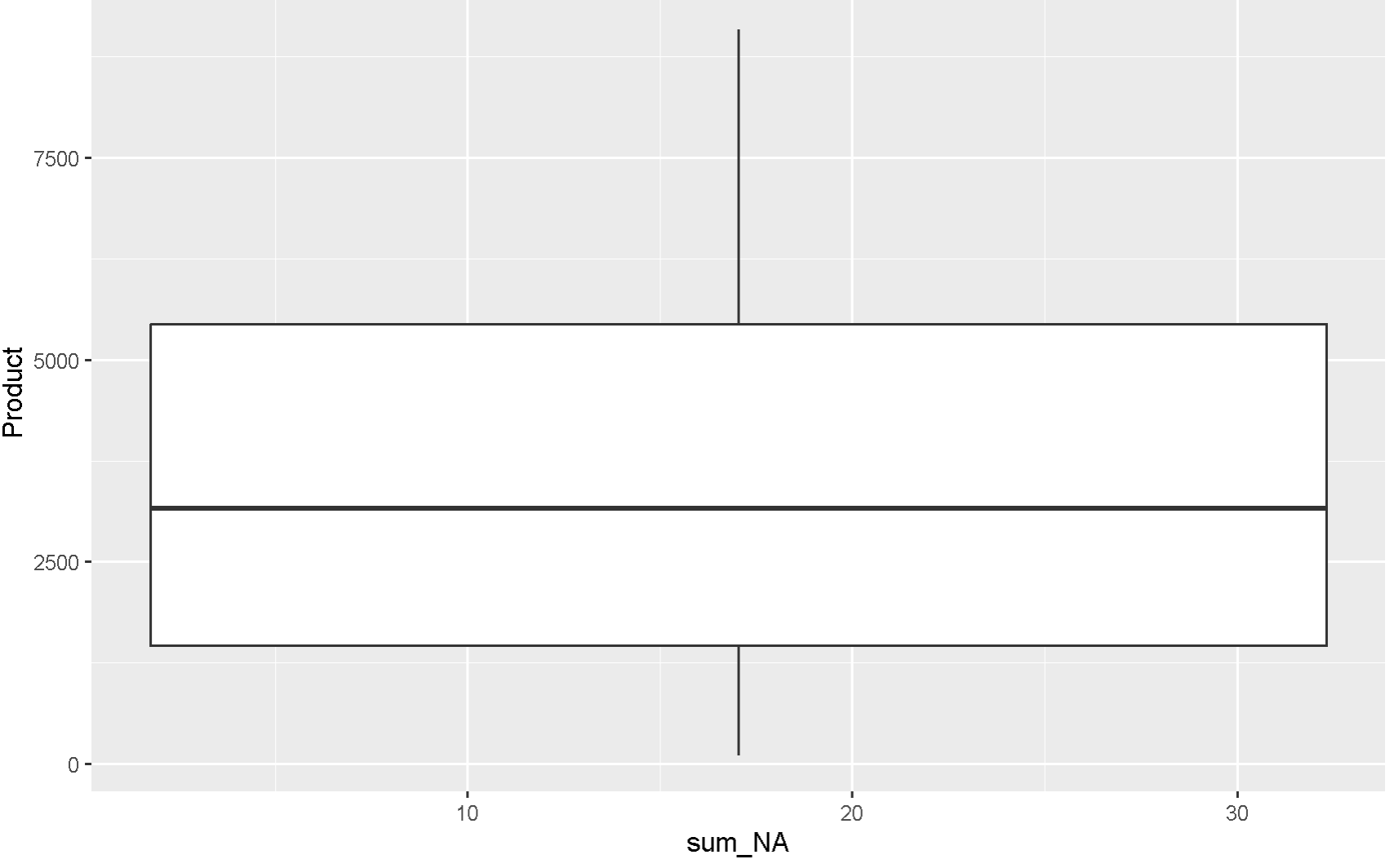


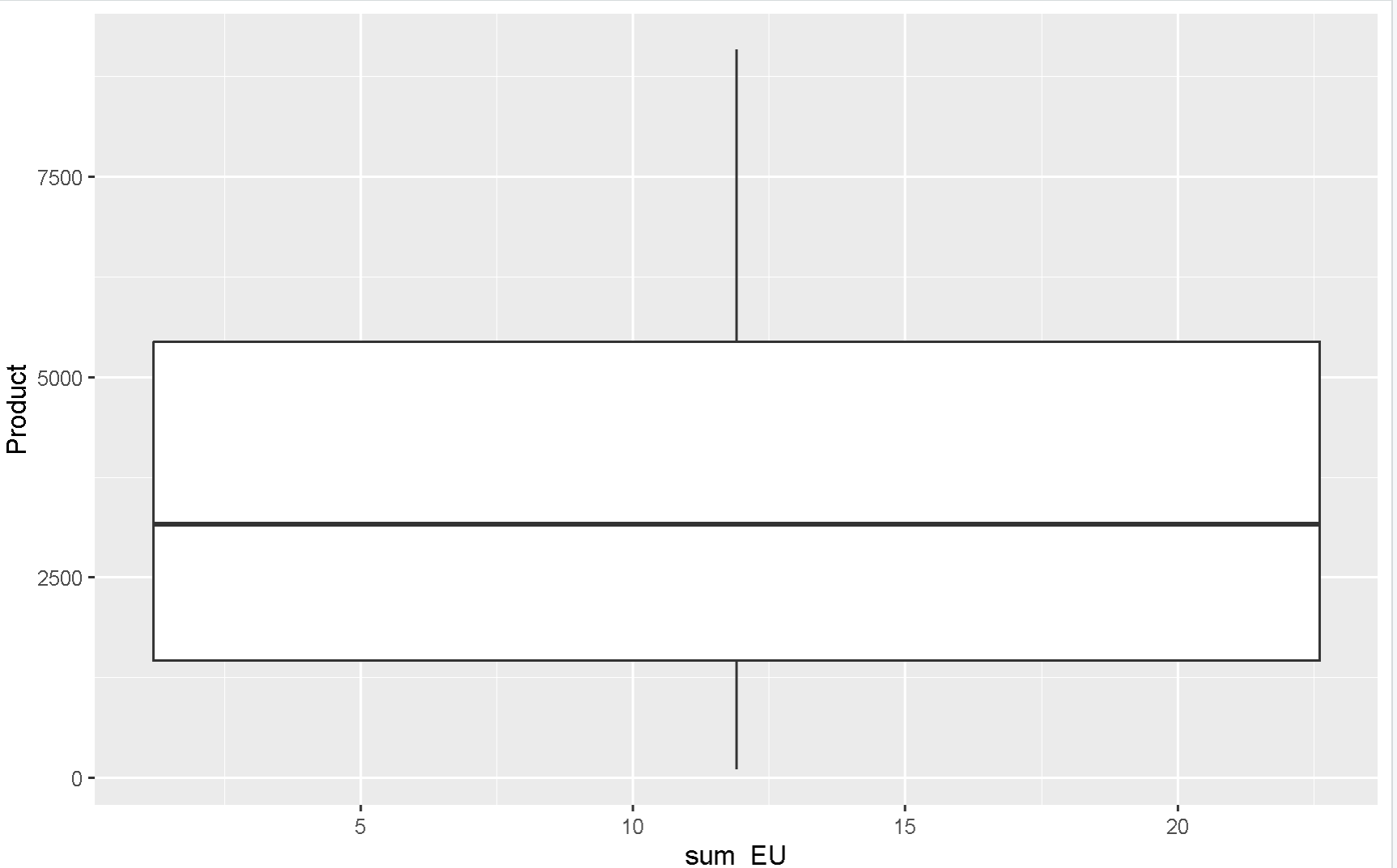


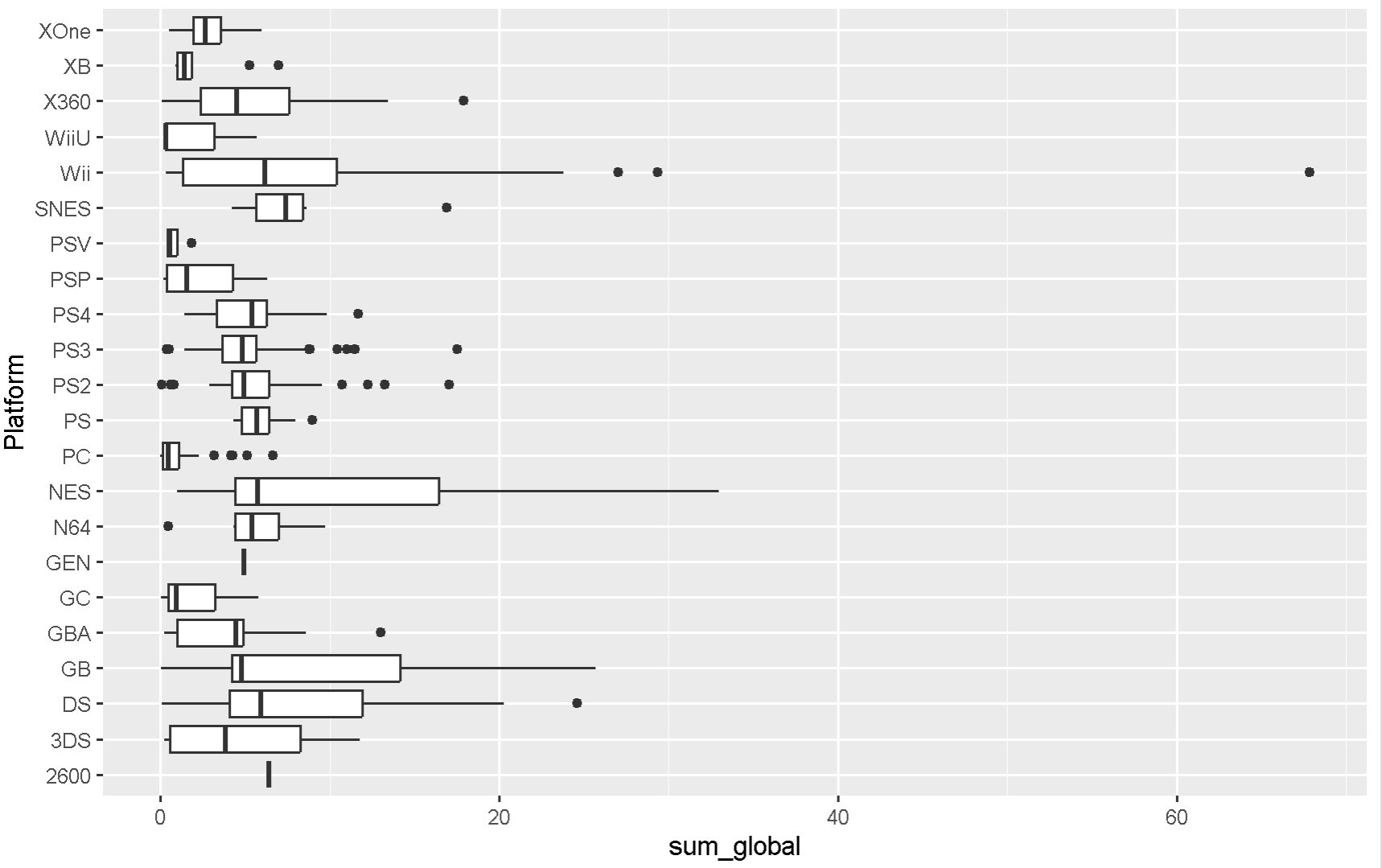


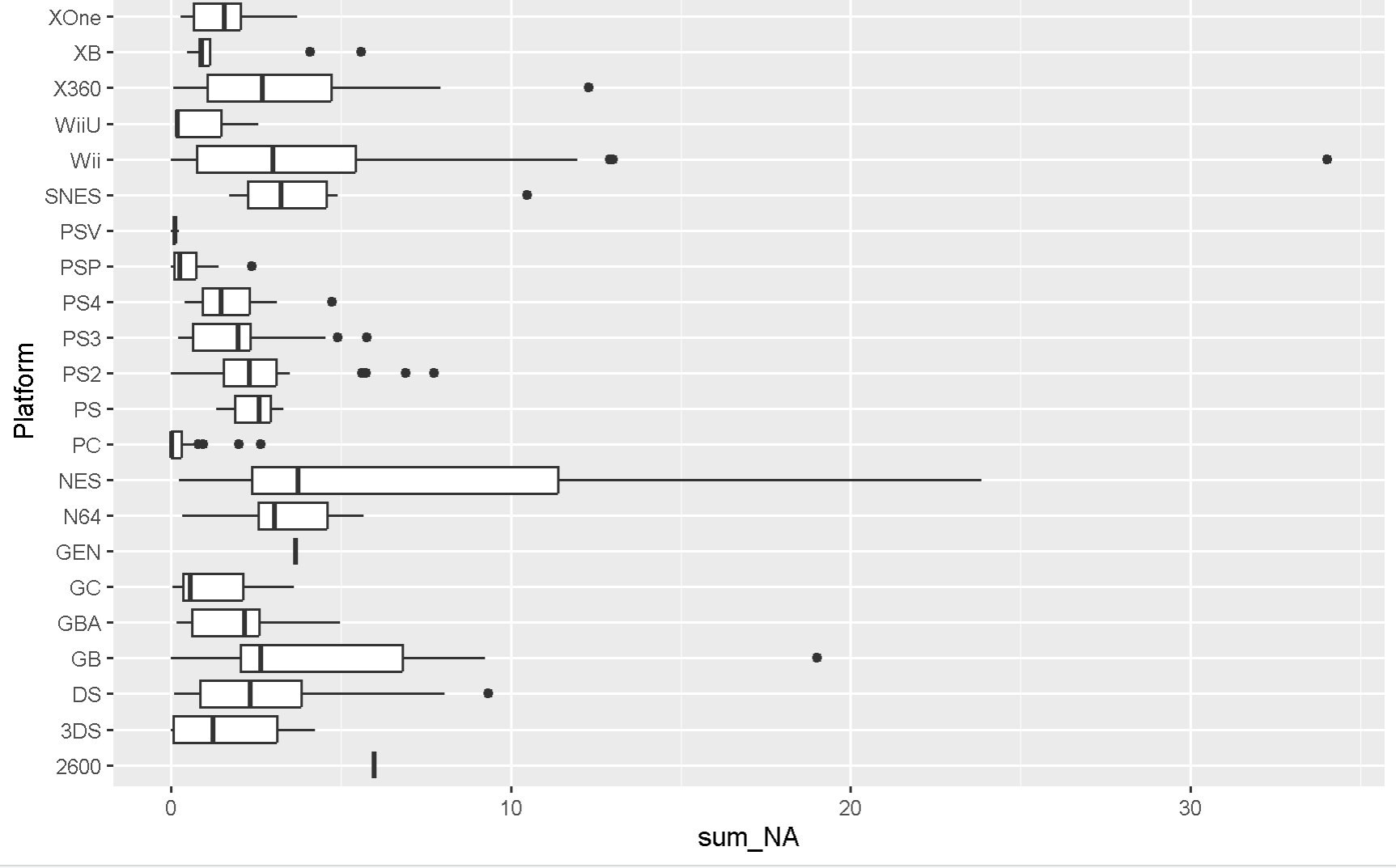


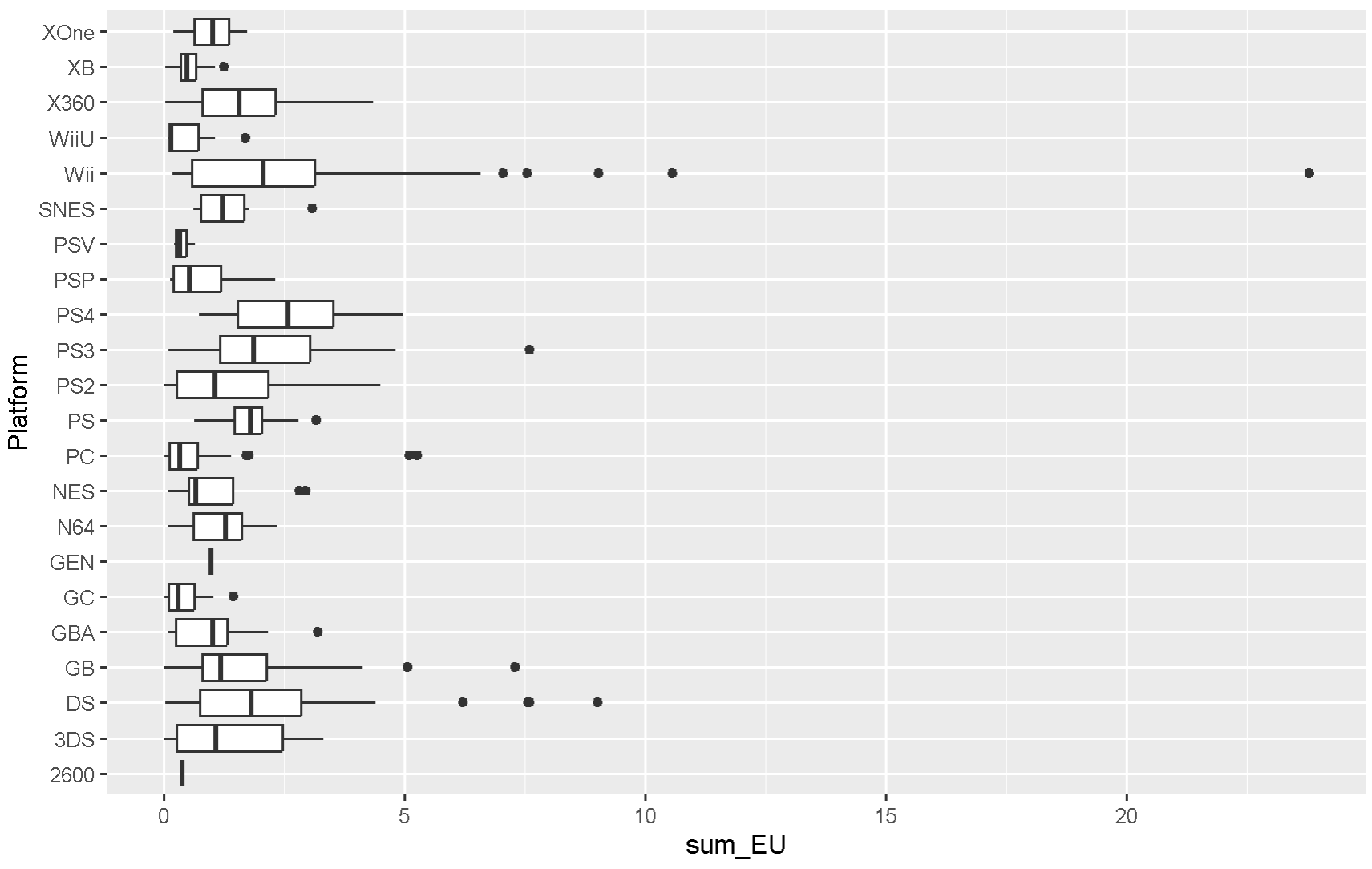












A computer code with numbers and letters

Description automatically generated

A graph of a normal q-q plot

Description automatically generated

A graph of a normal q-q plot

Description automatically generatedA screenshot of a computer code

Description automatically generated

A graph of a normal q-q plot

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer code

Description automatically generated