Exploratory Analysis for Module 4 Exercise

Hepatitis Case Data By Strain

Sophia Drewry + Amelia Foley

2021-09-16

# 1 Summary/Abstract

This document summarizes the results of our analysis from the Module 4 data wrangling exercise. Data was obtained on Hepatitis A, B, and C cases by week.

## 1.1 Questions/Hypotheses to be addressed

How do cases of Hepatitis A, B, and C change over time throughout the course of the ~60 week study period?

## 1.2 Data import and cleaning

Data import and cleaning was performed in the R Script “processingscript.R”. Data was imported from the National Notifiable Diseases Surveillance System via the CDC. Processing included renaming variables, assigning NA values, and refining the data set to include only a specific geographic area (the South).

## 1.3 Exploratory analysis

Table 1.1 shows a table summarizing the data.

Table 1.1: Data summary table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | area | week | hepA | hepB | hepC |
| Min. | 53 | 1 | 0 | 2 | 0 |
| 1st Qu. | character | 14 | 2 | 6 | 2 |
| Median | character | 27 | 3 | 8 | 3 |
| Mean | 53 | 27 | 3.43396226415094 | 8.94339622641509 | 3.60377358490566 |
| 3rd Qu. | character | 40 | 5 | 12 | 5 |
| Max. | character | 53 | 9 | 23 | 20 |

Figure 1.1 shows a boxplot figure produced by one of the R scripts. It appears that hepA and hepC exhibited similar case numbers, while hepB had a greater number of weekly cases throughout the study period.

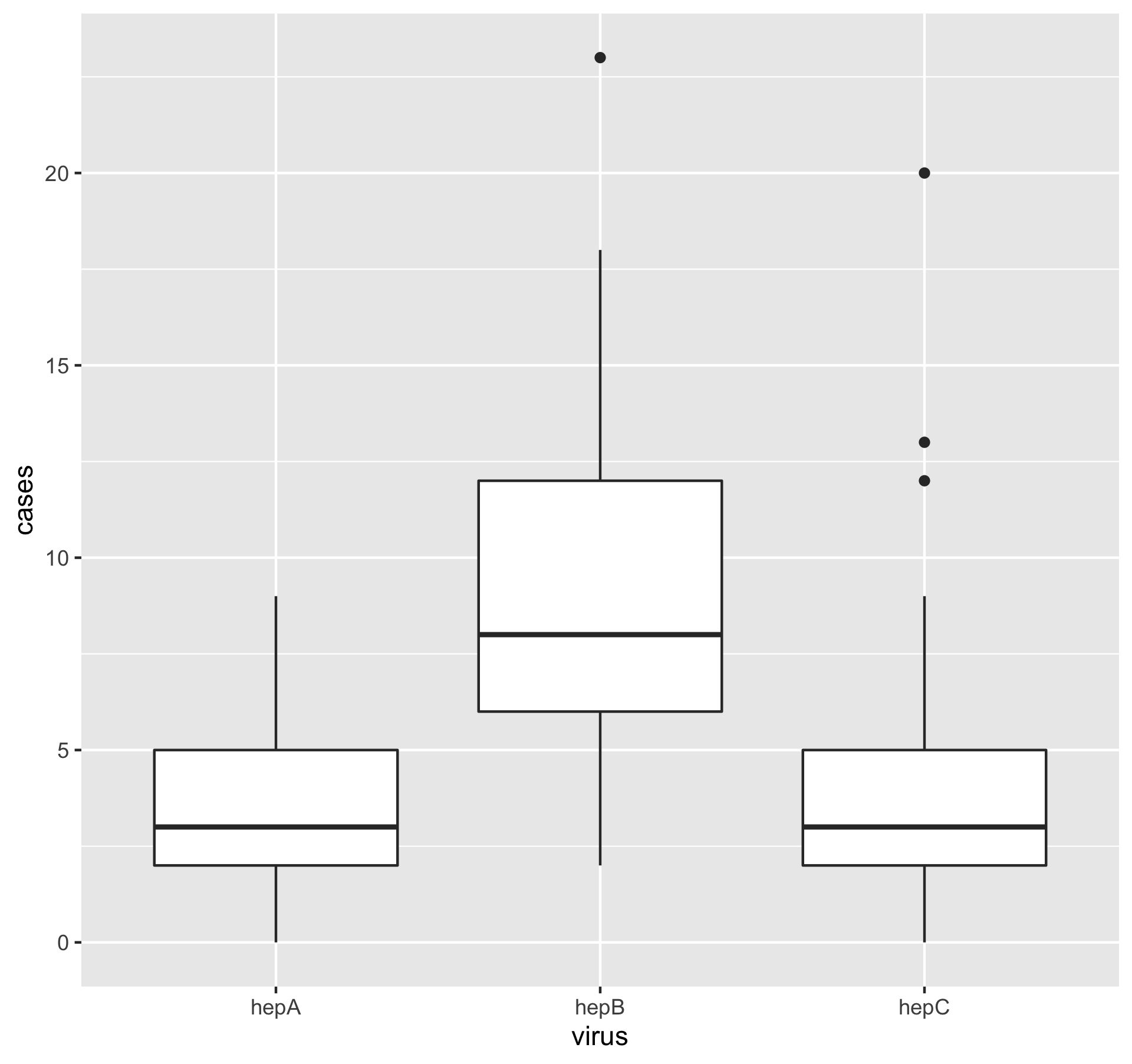


Figure 1.1: Analysis figure.

Figure 1.2 shows a scatterplot demonstrates the number of Hepatitis cases by strain over a period of weeks. No clear trends are immediately apparent by looking at this plot, though we see higher case numbers for hepB and similar case numbers for hepA and hepC.

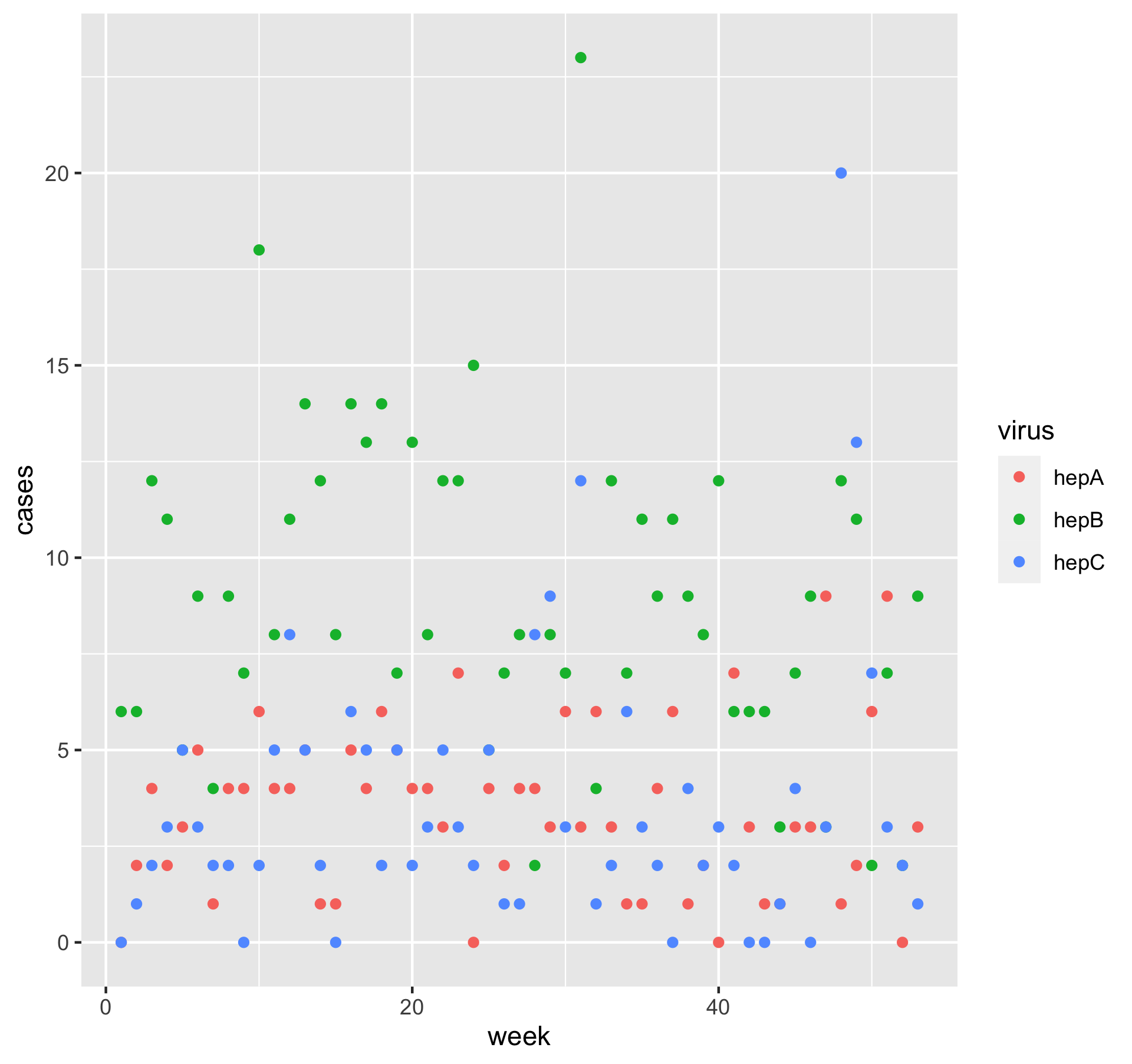


Figure 1.2: Analysis figure 2.

Figure 1.3 applies a best fit to the initial scatterplot. Based on this fit, hepB cases decrease throughout the study period, while hepC cases increase slightly and hepB cases remain constant.

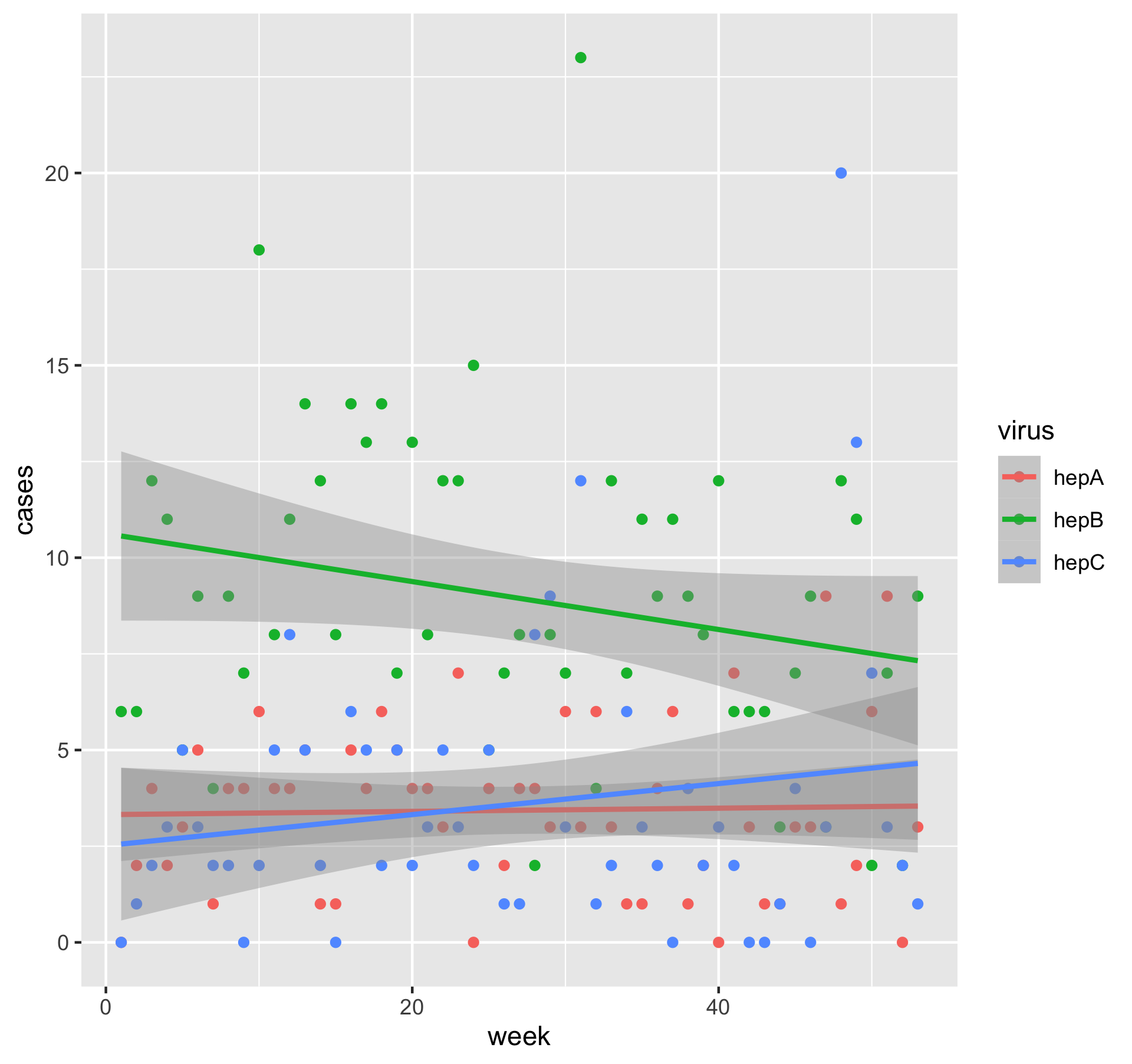


Figure 1.3: Analysis figure 3.

Figure 1.4 showcases the lines of best fit for each Hepatitis strain over time without the scatterplot points. There is a clear decrease in hepB cases, a slight increase in hepC cases, and a consistent number of hepA cases.

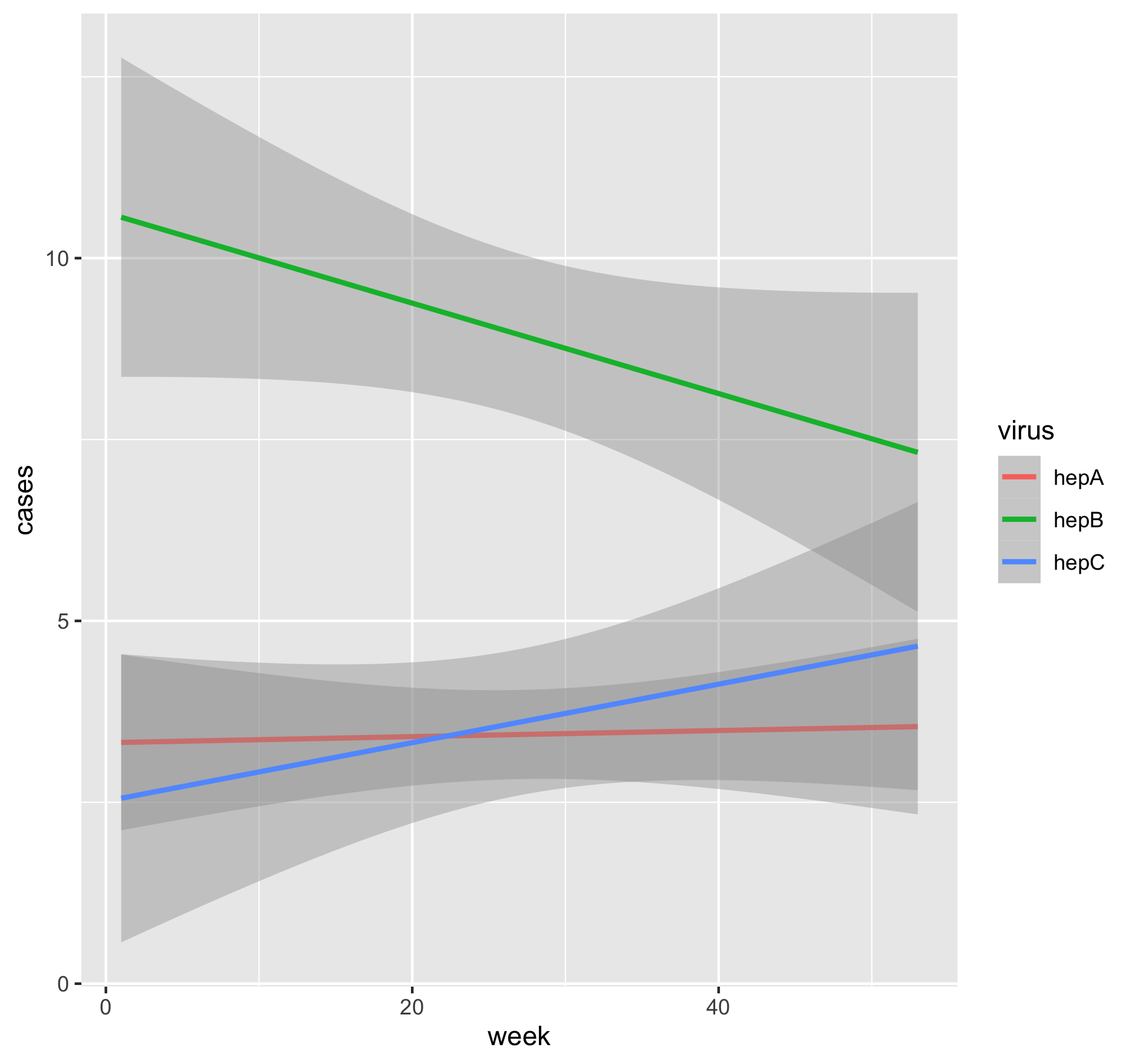


Figure 1.4: Analysis figure 4.