

Econometrics I: Assignment 1

Daisy Song

Date Due: 09/03/2025

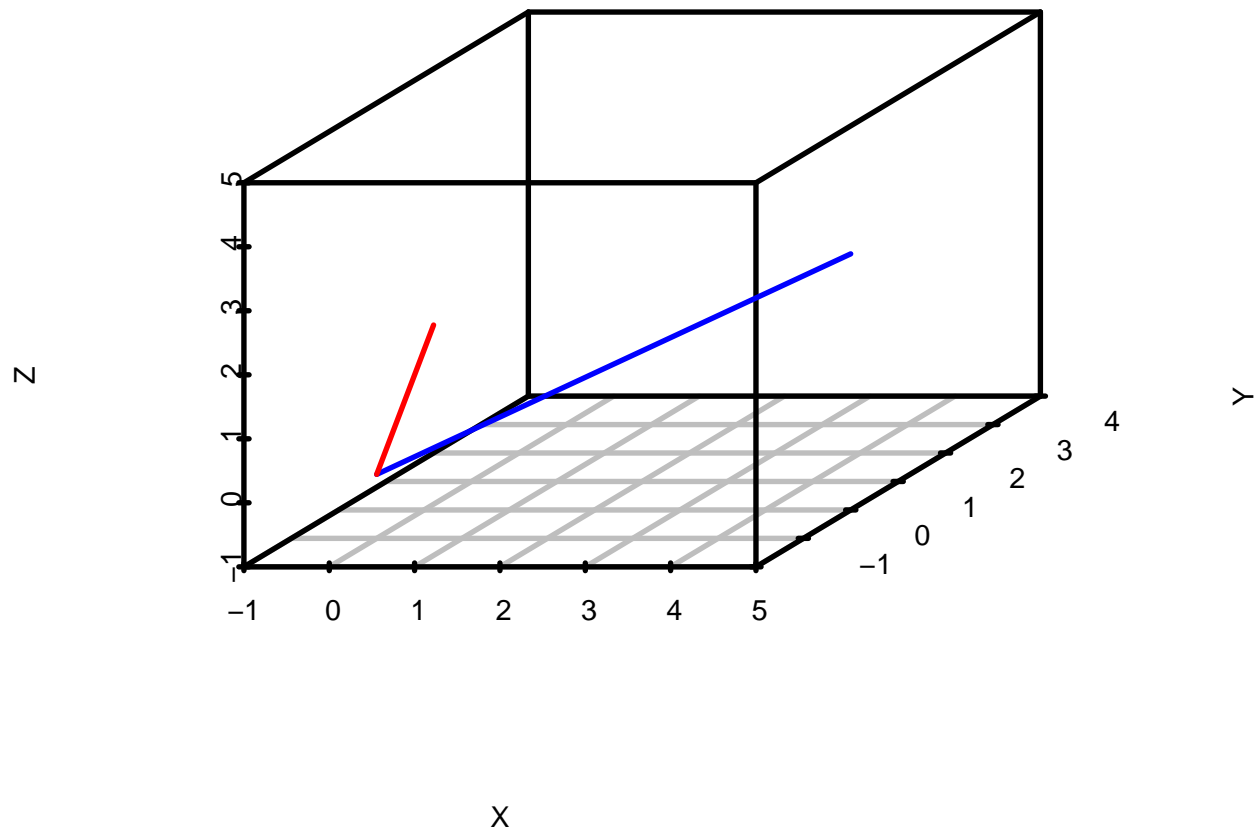
Problem 1(a)

Let $\mathbf{x}' = [5 \ 1 \ 3]$ and $\mathbf{y}' = [-1 \ 3 \ 1]$.

- Graph the two vectors.
- Find (i) the length of \mathbf{x} , (ii) the angle between \mathbf{x} and \mathbf{y} , and (iii) the projection \mathbf{y} on \mathbf{x} .
- Since $\bar{x} = 3$ and $\bar{y} = 1$, graph the centered vectors. (Hint: You can center the vector by subtracting the mean from each element: $\mathbf{x}^* = \mathbf{x} - \bar{x}\mathbf{1}$, where $\mathbf{1}' = [1 \ 1 \ 1]$.)
- In words describe how does recentering effect the relationship of the vectors?

Solution 1(a)

- Graph the two vectors.



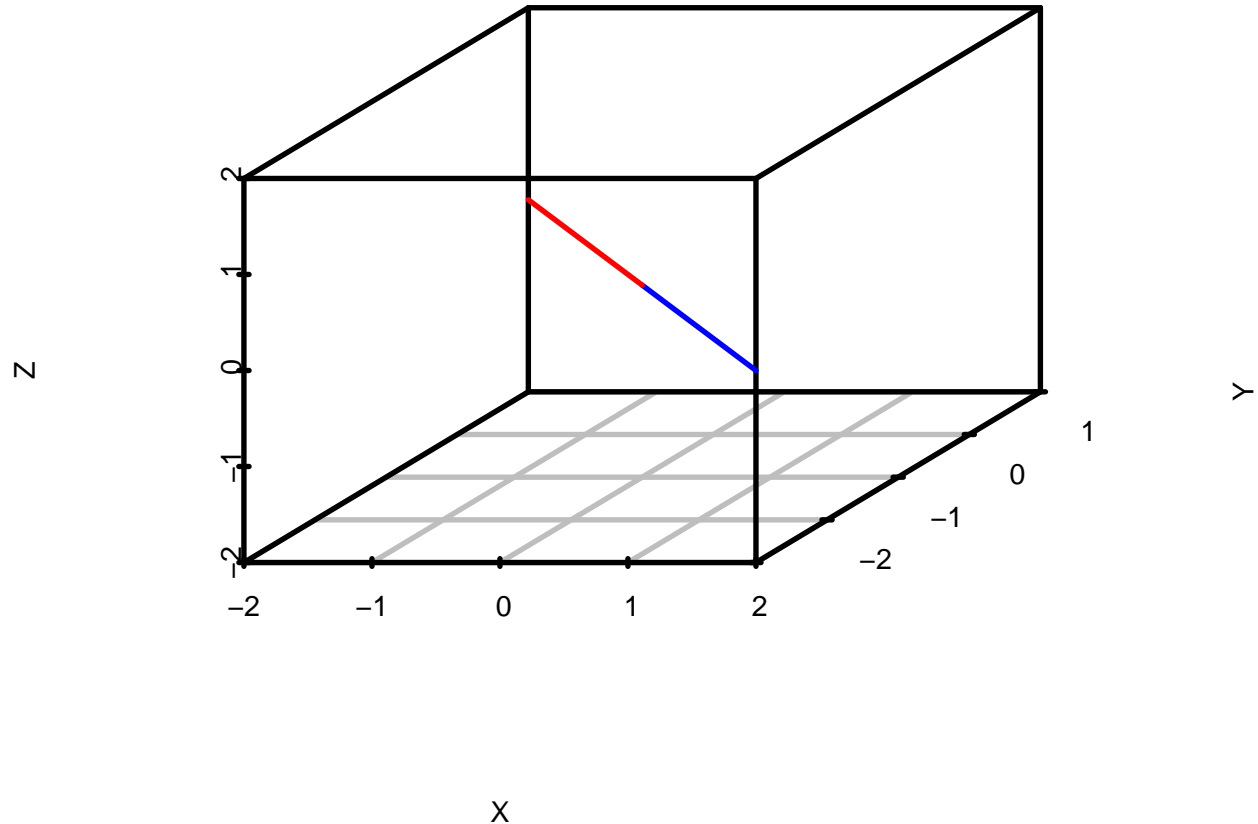
ii.

The length of x : $\sqrt{25 + 1 + 9} = \sqrt{35}$

The angle between x and y : $\cos(\theta) = \frac{x \cdot y}{\|x\| \|y\|} = \frac{-5+3+3}{\sqrt{35}\sqrt{11}} = \frac{1}{\sqrt{35 \cdot 11}} \implies \theta = \cos^{-1}\left(\frac{1}{\sqrt{35 \cdot 11}}\right)$

The projection of y onto x : $\text{proj}_x y = \frac{x \cdot y}{\|x\|^2} x = \frac{1}{35} \begin{bmatrix} 5 \\ 1 \\ 3 \end{bmatrix}$

iii.



Centered x is 2, -2, 0 Centered y is -2, 2, 0

iv.

Before centering, the relationship between vector x and y is affected by level of the value. Their angle is a number that is hard to interpret. After centering, the focus shifts to the variation within each vector, making it easier to compare their directions and relationships without the influence of their means, which significantly reduces the noise from constant terms but enhances correlation clarity. In our example, we can see that the two centered vectors are totally aligned with exactly opposite directions.

Problem 1(b)

Let $A = \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}$.

- Determine the eigenvalues and eigenvectors of A analytically.

- ii. Write the spectral decomposition of \mathbf{A} .
- iii. Find \mathbf{A}^{-1} .
- iv. Find the eigenvalues and eigenvectors of \mathbf{A}^{-1} .

Solution 1(b)

i. Determine the eigenvalues and eigenvectors of \mathbf{A} analytically.

$\mathbf{A} - \lambda \mathbf{I} = \begin{bmatrix} 9 - \lambda & -2 \\ -2 & 6 - \lambda \end{bmatrix}$ The characteristic polynomial is given by:

$$\det(\mathbf{A} - \lambda \mathbf{I}) = (9 - \lambda)(6 - \lambda) - (-2)(-2) = 0$$

Expanding this, we get:

$$\lambda^2 - 15\lambda + 50 = 0$$

Therefore λ are the roots of the characteristic polynomial, that is, $\lambda = 10, 5$.

ii. write the spectral decomposition of \mathbf{A} .

The spectral decomposition of a matrix \mathbf{A} is given by:

$$\mathbf{A} = \sum_{i=1}^n \lambda_i \mathbf{e}_i \mathbf{e}_i^T$$

where λ_i are the eigenvalues and \mathbf{e}_i are the corresponding eigenvectors.

When $\lambda = 5$, the eigenvector is $\mathbf{e}_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

When $\lambda = 10$, the eigenvector is $\mathbf{e}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

Thus $\mathbf{A} = 5\mathbf{e}_1 \mathbf{e}_1^T + 10\mathbf{e}_2 \mathbf{e}_2^T$.

iii. find \mathbf{A}^{-1}

Method 1: use the inverse equation for two by two matrix:

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \text{adj}(\mathbf{A})$$

Method 2: we can use the spectral decomposition:

Substituting the eigenvalues and eigenvectors, we get:

$$\mathbf{A}^{-1} = \frac{1}{5} \mathbf{e}_1 \mathbf{e}_1^T + \frac{1}{10} \mathbf{e}_2 \mathbf{e}_2^T$$

iv. find the eigenvalues and eigenvectors of \mathbf{A}^{-1} .

According to the properties of eigenvalues, the eigenvalues of \mathbf{A}^{-1} are given by $\frac{1}{\lambda_i}$, where λ_i are the eigenvalues of \mathbf{A} .

Thus, the eigenvalues of \mathbf{A}^{-1} are $\frac{1}{5}$ and $\frac{1}{10}$.

The eigenvectors remain the same:

When $\lambda = 5$, the eigenvector is $\mathbf{e}_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

When $\lambda = 10$, the eigenvector is $\mathbf{e}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

Problem 1(c)

Let S be a real symmetric $n \times n$ matrix. Show that if S^{-1} exists, then it is also symmetric.

Solution 1(c)

If S is symmetric, then $S^T = S$. If S is invertible, then $SS^{-1} = I$. Replace S with S^T in the equation:

$$S^T S^{-1} = I$$

Thus, $(S^T)^{-1} = S^{-1} = (S^{-1})^T$, which shows that S^{-1} is also symmetric.

Problem 1(d)

Show that the determinant of a square $p \times p$ matrix \mathbf{A} can be expressed as a product of its eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$. (Hint: use the spectral decomposition)

Solution 1(d)

Given that \mathbf{A} is a square matrix with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$, \mathbf{A} can be factored as

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$$

Taking the determinant of both sides, we have:

$$\det(\mathbf{A}) = \det(\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}) = \det(\mathbf{Q}) \det(\mathbf{\Lambda}) \det(\mathbf{Q}^{-1}) = \det(\mathbf{\Lambda})$$

Using the property of determinants, we can express the determinant as a product of the eigenvalues:

$$\det(\mathbf{A}) = \prod_{i=1}^p \lambda_i$$

Problem 1(e)

Find the minimum and maximum value of $4x_1^2 + 4x_2^2 + 6x_1x_2$ subject to the condition $x_1^2 + x_2^2 = 1$. (Hint: Recognize that the constraint is restricting solutions to a vector of unit length, and rewrite the quadratic equation in matrix form, and use spectral decomposition.)

Solution 1(e)

We can rewrite the quadratic form as:

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = 4x_1^2 + 4x_2^2 + 6x_1x_2$$

where

$$\mathbf{A} = \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}$$

We can find the eigenvalues and eigenvectors of \mathbf{A} to determine the extrema subject to the constraint $x_1^2 + x_2^2 = 1$.

The eigenvalues of \mathbf{A} are given by the characteristic polynomial:

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

Calculating this, we find the eigenvalues to be $\lambda_1 = 7$ and $\lambda_2 = 1$.

The corresponding eigenvectors can be found by solving the equation:

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{e} = 0$$

For $\lambda_1 = 7$, we find the eigenvector $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

For $\lambda_2 = 1$, we find the eigenvector $\mathbf{e}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Given that $x_1^2 + x_2^2 = 1$, the minimum value 1 occurs at the eigenvector corresponding to the smallest eigenvalue, which is \mathbf{e}_2 . The maximum value 7 occurs at the eigenvector corresponding to the largest eigenvalue, which is \mathbf{e}_1 .

Problem 1(f)

f. Answer the following questions using this covariance matrix: $\Sigma = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$ and the definition of

eigenvalues: $|\Sigma - \lambda \mathbf{I}| = 0$ and eigenvectors: $\Sigma \mathbf{e}_i = \lambda_i \mathbf{e}_i$.

- Derive the eigenvalues and eigenvectors of the covariance matrix from the definitions.
- Suppose $\rho = 0.5$ compute the eigenvalue and eigenvectors. What is the linear combination that would explain the most variation? What is the interpretation that you would give to this combination? What is the interpretation of the other two eigenvectors?

Solution 1(f)

i.

To find the eigenvalues and eigenvectors of the covariance matrix Σ , we start with the characteristic polynomial:

$$\det(\Sigma - \lambda \mathbf{I}) = 0$$

Substituting the expression for Σ , we have:

$$\det \left(\begin{bmatrix} 1-\lambda & \rho & \rho \\ \rho & 1-\lambda & \rho \\ \rho & \rho & 1-\lambda \end{bmatrix} \right) = 0$$

Calculating the determinant, we find the characteristic polynomial:

$$(1-\lambda)^3 - 3\rho^2(1-\lambda) + 2\rho^3 = 0$$

This cubic equation can be solved for λ to find the eigenvalues, which is $\lambda_1 = 1 + 2\rho$, $\lambda_2 = 1 - \rho$, $\lambda_3 = 1 - \rho$.

The eigenvector for $\lambda_1 = 1 + 2\rho$ is proportional to $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

The eigenvectors for $\lambda_2 = 1 - \rho$ (which has multiplicity 2) are any two orthogonal vectors in the subspace orthogonal to $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, for example: $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$ (or any orthogonal basis for this subspace).

Thus, the eigenvectors are:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

ii.

Substituting $\rho = 0.5$ into the characteristic polynomial, we can find the eigenvalues are $\lambda_1 = 2$, $\lambda_2 = 0.5$, $\lambda_3 = 0.5$. The largest eigenvalue 2 will correspond to the direction of maximum variance in the data.

The interpretation is that the first principal component (the direction of maximum variance) is a linear combination of all three original variables, indicating that they all contribute to this direction. The other two eigenvectors represent directions of lower variance, which are orthogonal to the first and to each other.

Problem 1(g)

[**Bonus**] Recall that a $n \times n$ matrix P is a projection matrix if it is symmetric and idempotent: $P^2 = P$. Show that P is a projection matrix if and only if it has the form

$$P = U \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} U^T$$

for some orthogonal matrix $U : n \times n$ and some $m \leq n$.

Solution 1(g)

Firstly, show that if P is a projection matrix, then it can be expressed in the desired form. Since P is symmetric and idempotent, we can perform an eigenvalue decomposition:

$$P = Q\Lambda Q^T$$

where Q is an orthogonal matrix whose columns are the eigenvectors of P , and Λ is a diagonal matrix containing the eigenvalues of P .

Since P is idempotent, the eigenvalues must be either 0 or 1. Let m be the number of 1 eigenvalues (the rank of P), then we can write:

$$\Lambda = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}$$

Thus, we have:

$$P = Q \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} Q^T$$

Now, we need to show the converse: if P has the form

$$P = U \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} U^T$$

for some orthogonal matrix U and some $m \leq n$, then P is a projection matrix.

To show this, we need to verify that P is symmetric and idempotent.

1. **Symmetry:** Since U is orthogonal, we have:

$$P^T = \left(U \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} U^T \right)^T = U \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} U^T = P$$

Thus, P is symmetric.

2. **Idempotence:** We need to show that $P^2 = P$:

$$P^2 = \left(U \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} U^T \right) \left(U \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} U^T \right) = U \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} U^T$$

Thus, $P^2 = P$, and P is idempotent.

Since P is both symmetric and idempotent, it is a projection matrix.

Problem 2

In this question we want to examine distributions with the help of R. In particular, we want to examine the (i) skewness, (ii) symmetry of a distribution, (iii) outliers and (iv) non-normality of the data sets Leinhardt. The dataset can be found in the package `car`, so first install and load the package `car`. To load the Leinhardt data set use the command `data(Leinhardt)`.

Conduct an exploratory analysis of the Leinhardt dataset.

- Based upon the description of the dataset from `?Leinhardt` what relationships do you expect exist? What do the variables mean? How many missing values?
- Examine the variable `infant` in the data set `Leinhardt` with respect to (i)-(iv) above (in the description).
- Create a single graphic to visualize the relationships amongst the variables. Comment on your findings. (Do not build a model)

```
# R Hints
data("Leinhardt")           # To read in data
summary(Leinhardt)          # To get an overview
boxplot(Leinhardt)           # Boxplot of the data

##      income      infant      region      oil
## Min.   : 50.0   Min.    : 9.60   Africa  :34   no  :96
## 1st Qu.:123.0   1st Qu.:26.20   Americas:23  yes: 9
## Median :334.0   Median :60.60   Asia     :30
## Mean   :998.1   Mean    :89.05   Europe  :18
## 3rd Qu.:1191.0   3rd Qu.:129.40
## Max.   :5596.0   Max.    :650.00
##                NA's    :4
```

Solution 2(a)

Base on the summary statistics, I expect that there is a positive correlation between the variables. For example, countries with higher income per capita are likely to have lower infant mortality rates. The variable `infant` represents the infant mortality rate per 1000 live births, which is expected to be negatively correlated with income. The variable `income` represents the income per capita in US dollars. The variable `region` represents A factor with levels: Africa; Americas; Asia, Asia and Oceania; Europe. The variable `oil` stands for whether the country is an oil-exporting country or not. There are 4 missing values for infant while no missing values for other variables.

Solution 2(b)

```
Leinhardt$infant           # To access the variable infant in the data set Leinhardt

##      [1] 26.7 23.7 17.0 16.8 13.5 10.1 12.9 20.4 17.8 25.7 11.7 11.6
##      [13] 16.2 11.3 44.8 71.5  9.6 12.8 17.5 17.6 86.3 78.5 125.0  NA
##      [25] 28.1 300.0 58.0 650.0 51.7 59.6 170.0 78.0 62.8 54.4 48.8 27.8
##      [37] 79.1 22.1 26.2 13.6 32.0 60.9 46.0 34.1 65.1 20.4 15.1 19.1
##      [49] 26.2 76.3 40.4 43.3 259.0 60.4 137.0 180.0 114.0 58.2 63.7 39.3
##      [61] 138.0 21.3 58.0 159.2 149.0 10.2 38.6 67.9 21.7 27.0 153.0 100.0
##      [73] 400.0 124.3 200.0 150.0 100.0 190.0 160.0 109.6 84.2 216.0  NA 60.6
##      [85] 55.0  NA 102.0 148.3 120.0 187.0  NA 200.0 124.3 132.9 170.0 158.0
##      [97] 45.1 129.4 162.5 127.0 160.0 180.0 80.0 50.0 104.0

infant <- Leinhardt$infant  # Saves the variable income to another object
summary(infant)             # For an overview

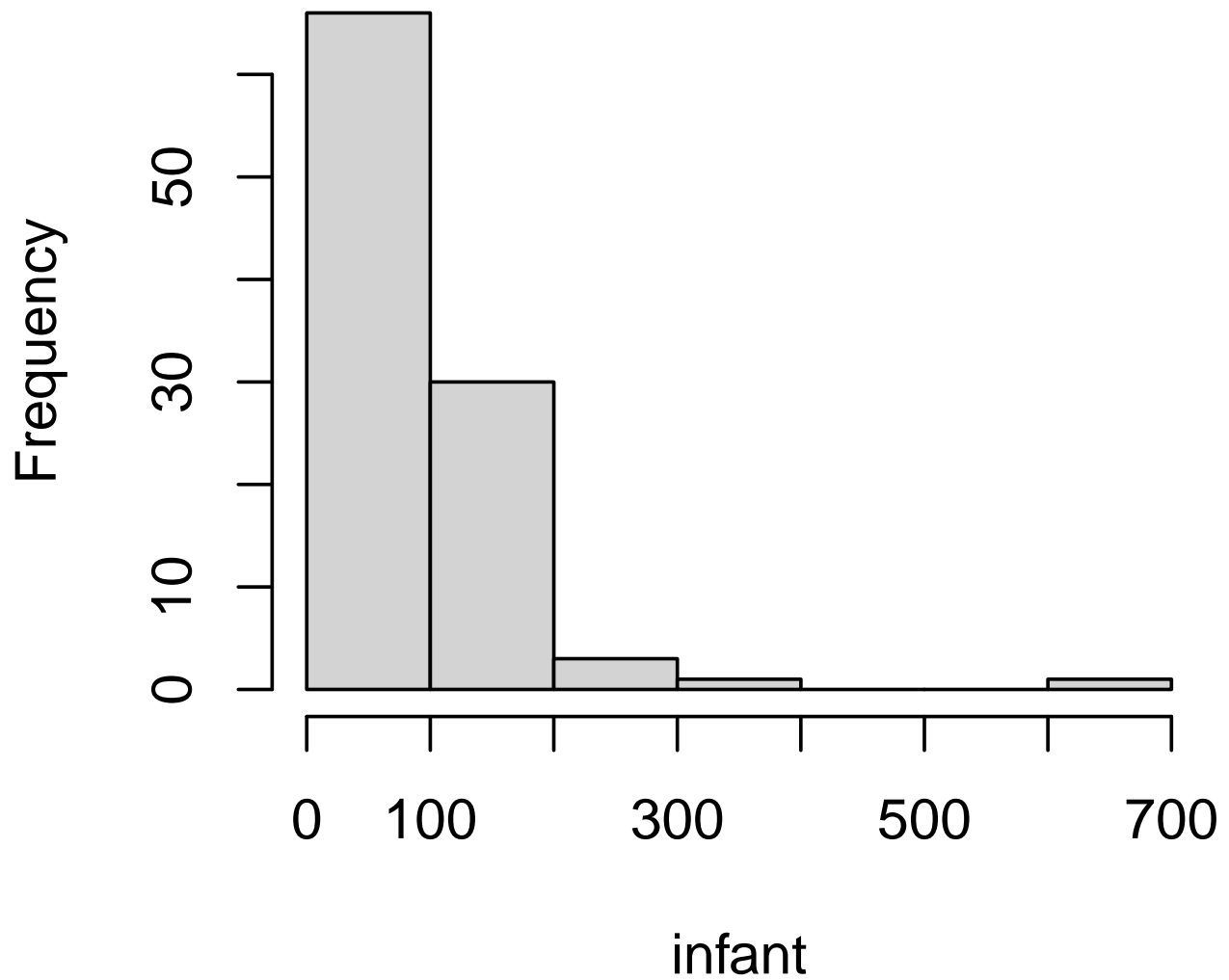
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##      9.60  26.20   60.60   89.05 129.40  650.00      4
```



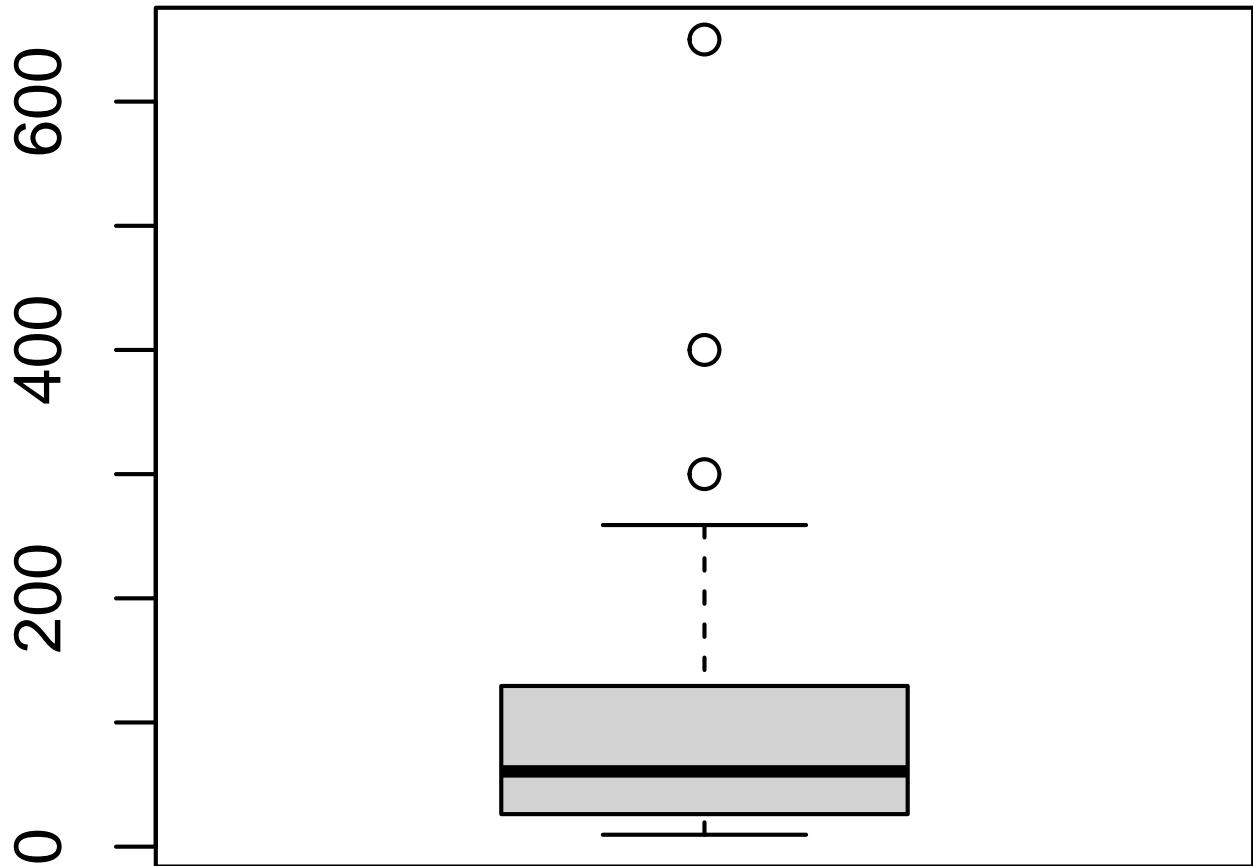
```
hist(infant)
```

```
# Draws a histogram
```

Histogram of infant



```
boxplot(infant)
```



Solution 2(c)

```
data("Leinhardt")
ggplot(
  Leinhardt,
  aes(
    x = income,
    y = infant,
    color = region,
    shape = oil
  )
) +
  geom_point(
    size = 2,
    alpha = 0.7
  ) +
  labs(
    title = "Income vs Infant Mortality Rate",
    x = "Income per Capita",
    y = "Infant Mortality Rate"
  ) +
  theme_minimal()
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Income vs Infant Mortality Rate

