# Basics in R – Assignment Answers

1. Open up R Studio and set your working directory to the folder you will be working from (the one that contains the data files you want to use).

In the window on the lower right corner – under Files – find the folder in the window – and then under More > Set As Working Directory

1. Save your R script into this folder. Include comments throughout your script to explain what each command is doing.

> save.image(“BasicsAssignments.R”)

or

File > Save As

1. What is one way that you can find help if you forget how to use the subset function?

> help(subset)

or

> ?subset

1. Using R as a calculator, what is the sum of 15 and 32?

> 15+32

[1] 47

5) Using R as a calculator, what is the product of 34 and 543?

> 34\*543

[1] 18462

6) Using R as a calculator, what is the value of 24?

> 2^4

[1] 16

7) Create a vector of the numbers 3,8,4,2 and give it the name A.

> A <- c(3,8,4,2)

> A

[1] 3 8 4 2

8) Create a vector of the characters monkey, rat, snake, and give it the name B.

> B <- c("monkey", "rat", "snake")

> B

[1] "monkey" "rat" "snake"

9) Create a vector of the sequence of numbers between 89 and 143, and give it the name C.

> C <- seq(89,143)

> C

[1] 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103

[16] 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118

[31] 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133

[46] 134 135 136 137 138 139 140 141 142 143

10) Create a vector containing 15 of the number 3 – use the rep command, and give it the name D.

> D <- rep(3,15)

> D

[1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3

11) What is the 3rd element in the vector A?

> A[3]

[1] 4

12) What are the last two elements of the vector B?

> B[2:3]

[1] "rat" "snake"

13) What is the 4th and 30th element in the vector C?

> C[c(4,30)]

[1] 92 118

14) What is the sum of all of the elements in the vector D?

> sum(D)

[1] 45

15) Add the character “bird” to the end of vector B.

> B[4] <- "bird"

> B

[1] "monkey" "rat" "snake" "bird"

16) Change the first element in vector D to the number 5.

> D[1] <- 5

> D

[1] 5 3 3 3 3 3 3 3 3 3 3 3 3 3 3

17) Multiple vector A by 3 and save the results as vector A2.

> A2 <- A\*3

> A2

[1] 9 24 12 6

18) Add 5 to every element of vector C and save the results as vector C2.

> C2 <- C+5

> C2

[1] 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108

[16] 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123

[31] 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138

[46] 139 140 141 142 143 144 145 146 147 148

1. Import the ‘lionCrater’ file into R and save the dataframe as ‘Lion”. This file gives the population numbers for lions in the Ngorongoro Crater in Tanzania.

> Lion <- read.csv("lionCrater.csv", header = T)

1. Import the ‘blackRhinoCrater’ file into R and save the dataframe as “Rhino”. This file gives the population numbers for black rhinos in the Ngorongoro Crater in Tanzania.

> Rhino <- read.csv("blackRhinoCrater.csv", header = T)

1. Display the structure of the Lion data frame. What is included? How many observations and how many variables? What are the variable names?

> str(Lion)

'data.frame': 30 obs. of 2 variables:

$ Year : int 1963 1964 1965 1966 1967 1968 1968 1970 1971 1972 ...

$ Population: int 13 12 31 25 39 46 60 61 53 69 ...

There are 30 observations in the data frame of 2 variables. The variables are year and population.

1. What are three ways to display the data in column 1 of the Lion data frame?

> Lion[1]

> Lion["Year"]

> Lion$Year

1. How do you access the 5th row of data in the Lion data frame?

> Lion[5,]

Year Population

5 1967 39

24) Display the first 6 rows of the Rhino data frame.

> head(Rhino)

Year Population

1. 1980 20
2. 1981 21
3. 1982 20
4. 1983 15
5. 1984 10
6. 1985 11

25) Display the last 6 rows of the Rhino data frame.

> tail(Rhino)

Year Population

1. 1990 11
2. 1991 12
3. 1992 14
4. 1993 15
5. 1994 16
6. 1995 16

26) How many rows are in the Rhino data frame?

> nrow(Rhino)

[1] 16

27) How many columns are in the Rhino data frame?

> ncol(Rhino)

[1] 2

28) Look at the Year column for both Lions and Rhinos. Which years overlap? Create two new data frames Lion2 and Rhino2 that only contain the overlapping years.

1980-1992 are the overlapping years.

> Lion2 <- subset(Lion, Lion$Year > 1979)

> Rhino2 <- subset(Rhino, Rhino$Year < 1993)

29) Create a new data frame called Lion3 that only contains the years when the population is over 50 individuals. How many rows are in this new data frame?

> Lion3 <- subset(Lion, Lion$Population > 50)

> nrow(Lion3)

[1] 23

30) Create a new data frame called Rhino3 that only contains the years when the population of rhinos is less than 12 individuals. How many rows are in this new data frame?

> Rhino3 <- subset(Rhino, Rhino$Population < 12)

> nrow(Rhino3)

[1] 6

# Descriptive Statistics/Plotting – Assignment Answers

1. Import the Lion and Black Rhino files into R that we were using in the last assignment (‘lionCrater.csv’ and ‘blackRhinoCrater.csv’).

> Lion <- read.csv("lionCrater.csv", header = T)

> Rhino <- read.csv("blackRhinoCrater.csv", header = T)

1. What is the minimum and maximum population size for lions?

> min(Lion$Population)

[1] 12

> max(Lion$Population)

[1] 123

3) What is the mean, median, and standard deviation of the population size of lions?

> mean(Lion$Population)

[1] 69.2

> median(Lion$Population)

[1] 72.5

> sd(Lion$Population)

[1] 27.80523

4) What is the minimum and maximum population size for black rhinos?

> min(Rhino$Population)

[1] 10

> max(Rhino$Population)

[1] 21

5) What are the quantile (0%,25%,50%,75%,100%) population sizes for black rhinos?

> quantile(Rhino$Population)

0% 25% 50% 75% 100%

1. 11 14 16 21

6) Use the summary command to find the descriptive statistics of population size for black rhinos. What is the mean and median?

> summary(Rhino$Population)

Min. 1st Qu. Median Mean 3rd Qu. Max.

10.00 11.00 14.00 14.12 16.00 21.00

The mean is 14.12 rhinos and the median is 14 rhinos.

7) What is the variance and standard deviation of population size for black rhinos?

> var(Rhino$Population)

[1] 13.98333

> sd(Rhino$Population)

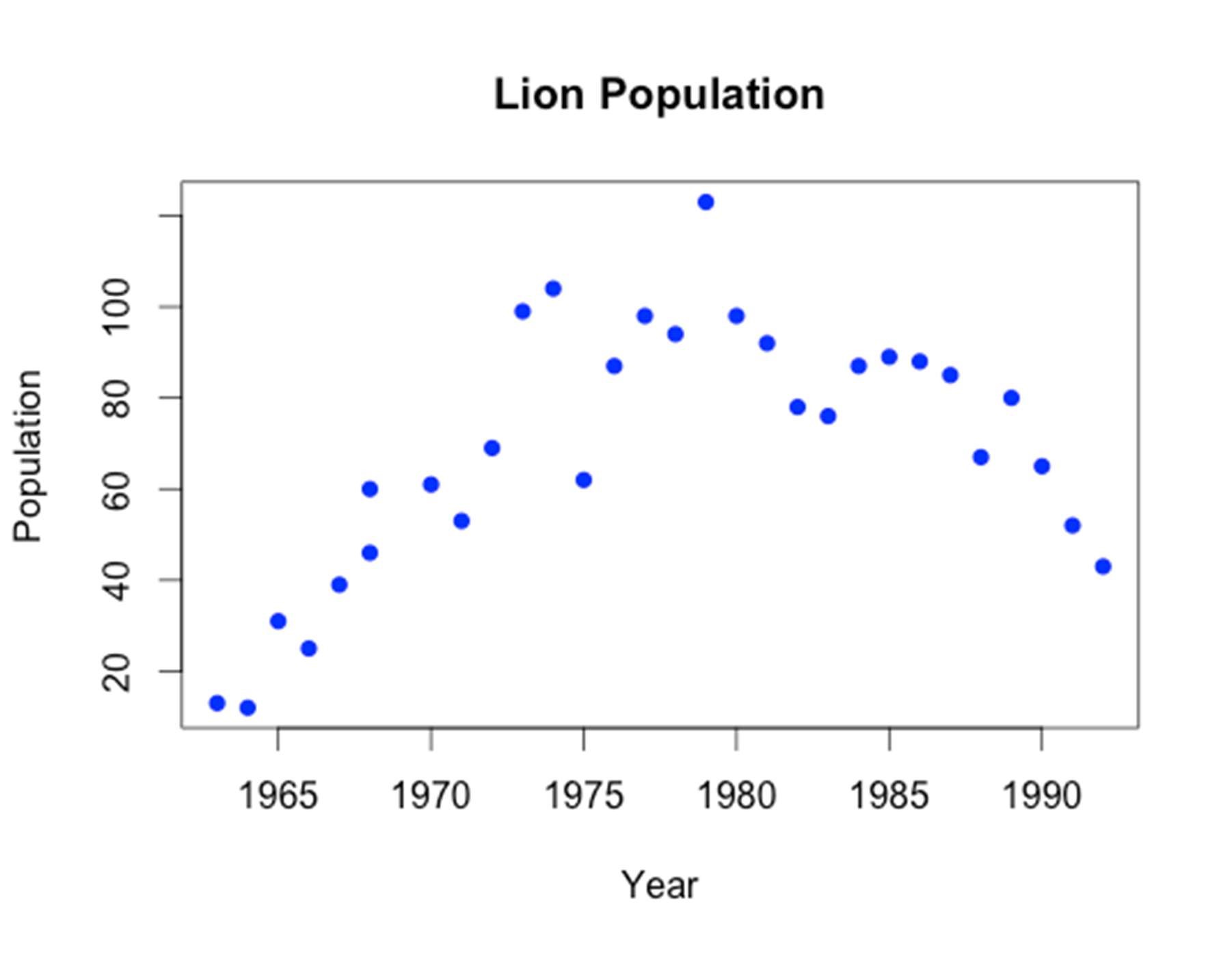
[1] 3.73943

1. Create a plot of the lion population size with the year on the x-axis. Include a title, axis labels, and make the points blue. Can you figure out how to change the open circle points to another shape? Try and change them to closed circles. Is the population increasing or decreasing? What year is the minimum and maximum population sizes?

> plot(Lion$Year, Lion$Population, main = "Lion Population", xlab = "Year", ylab =

"Population", col = "blue", pch = 16)

The population is increasing until 1979 and then decreasing. The minimum value is at 1964 and the maximum value is at 1979.

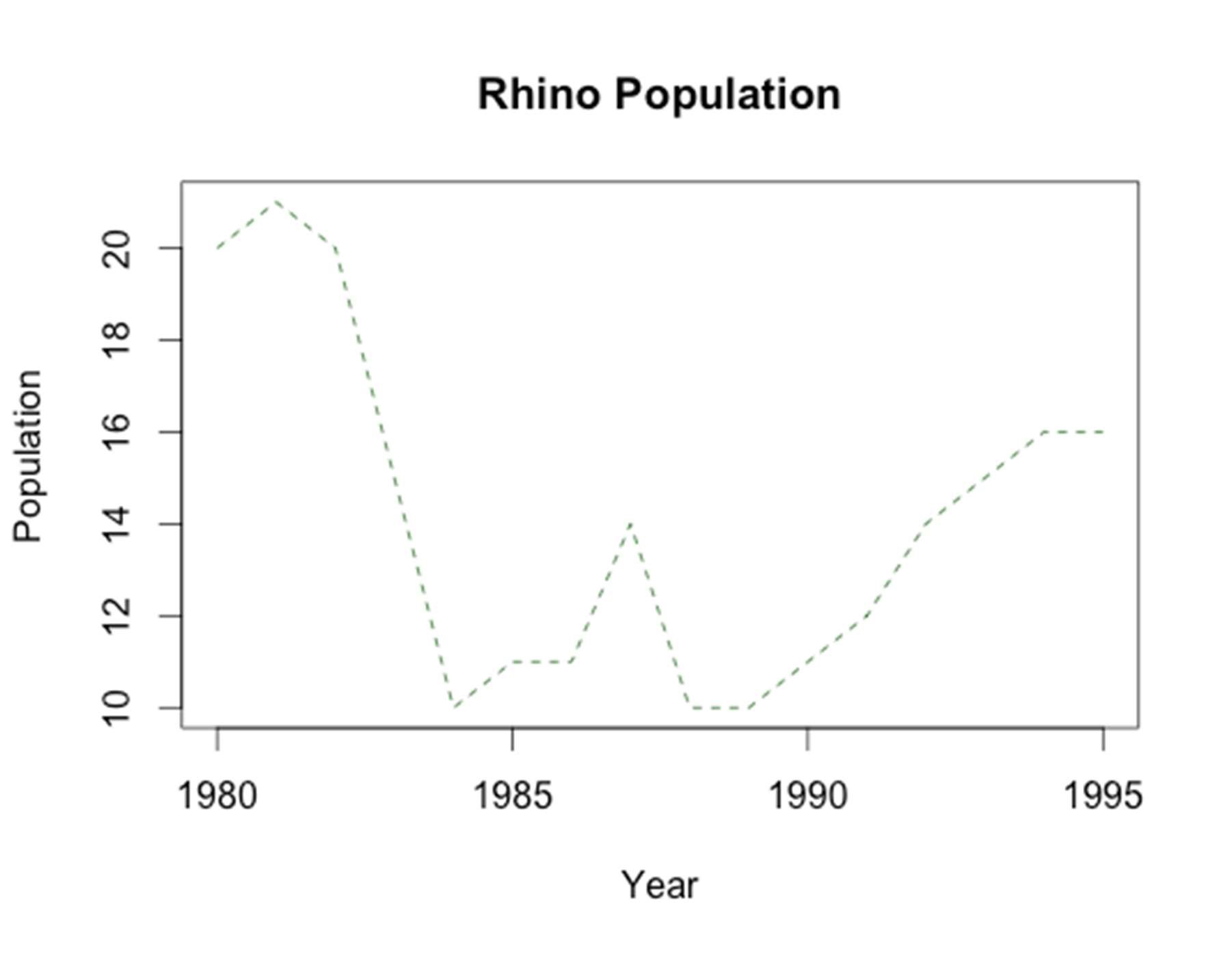


1. Create a plot of the black rhino population size with the year on the x-axis. This time make it a line plot. Include a title, axis labels, and the make the line dark green. Can you figure out how to change the line type to a dotted line? Is the population increasing or decreasing? What year is the minimum and maximum population sizes?

> plot(Rhino$Year, Rhino$Population, main = "Rhino Population", xlab = "Year", ylab =

"Population", col = "darkgreen", type = "l", lty = 2)

The population increases to 1981 then decreases to 1984 then increases to 1987 then decreases to 1988 and then increases to 1995. The minimum population size is at year 1984, 1988, and 1989, and the maximum population size is at 1981.

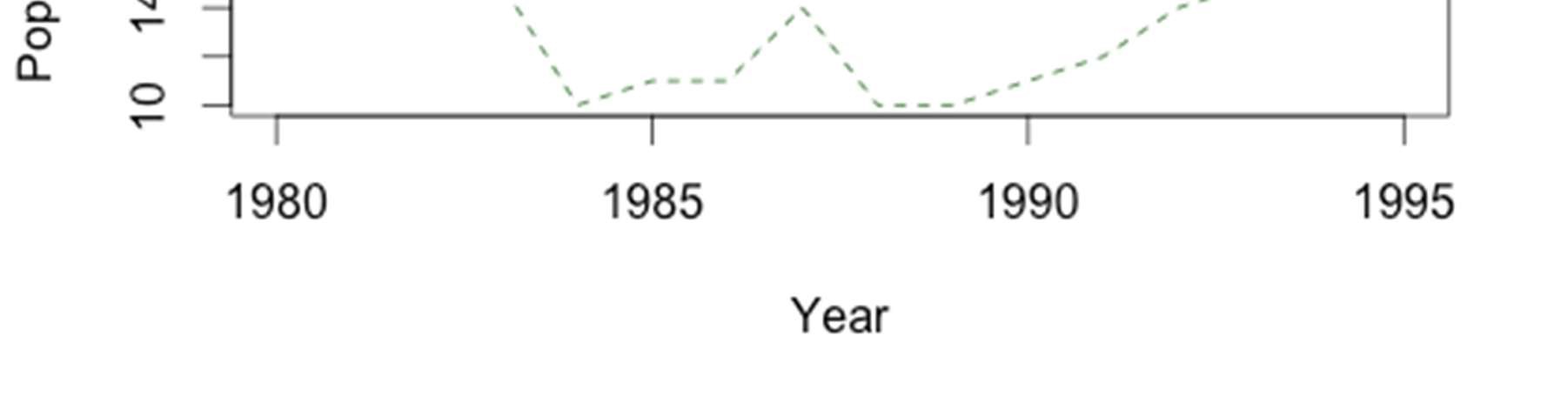
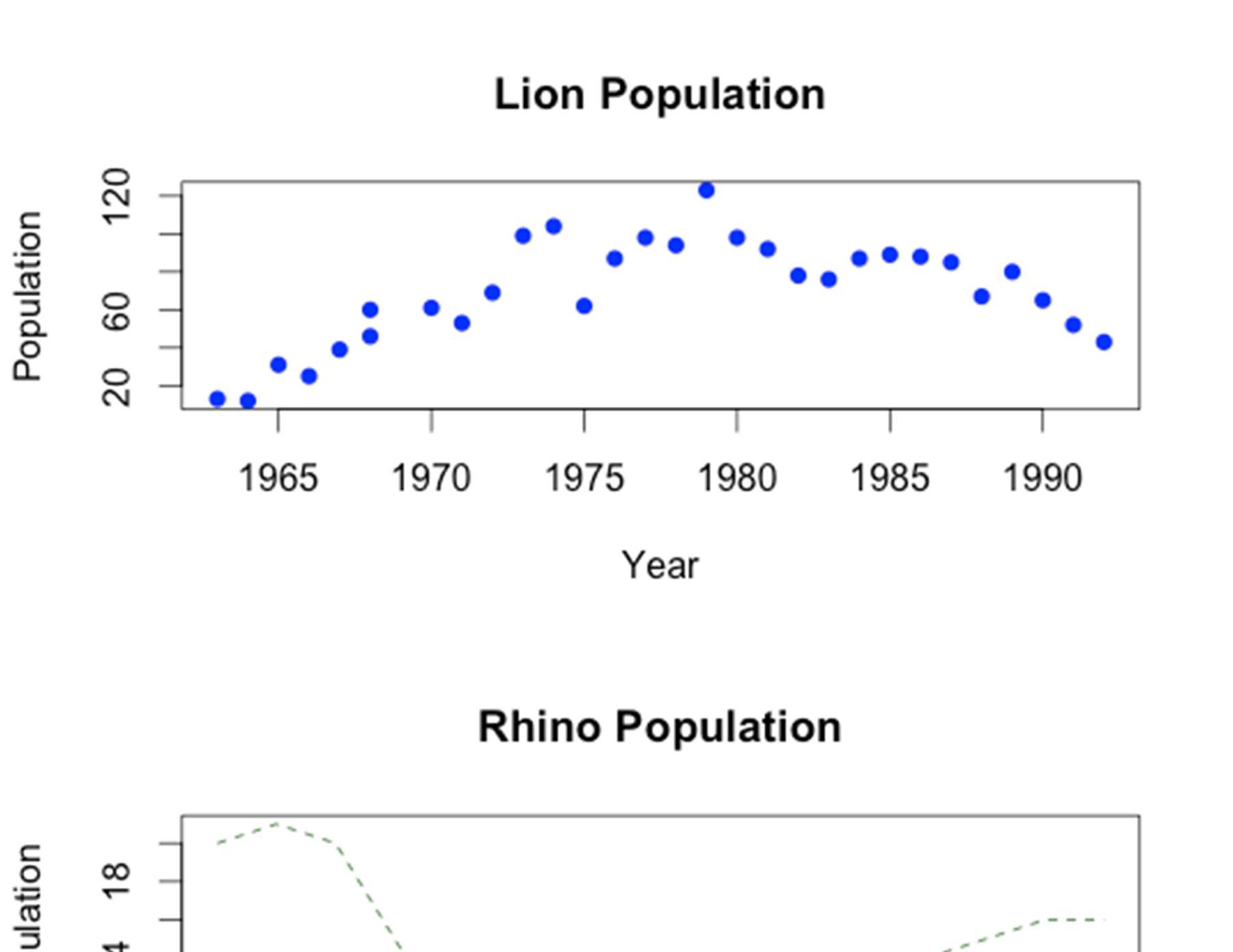


1. Put the two plots that you just made onto one page with one column and two rows.

> par(mfrow = c(2,1))

> plot(Lion$Year, Lion$Population, main = "Lion Population", xlab = "Year", ylab = "Population", col = "blue", pch = 16)

> plot(Rhino$Year, Rhino$Population, main = "Rhino Population", xlab = "Year", ylab = "Population", col = "darkgreen", type = "l", lty = 2)



1. Create two new data frames Lion2 and Rhino2 that contain the subset of years that overlap between the two datasets (like we did in the last assignment). These new datasets will be used for the next plot.

> Lion2 <- subset(Lion, Lion$Year >1979)

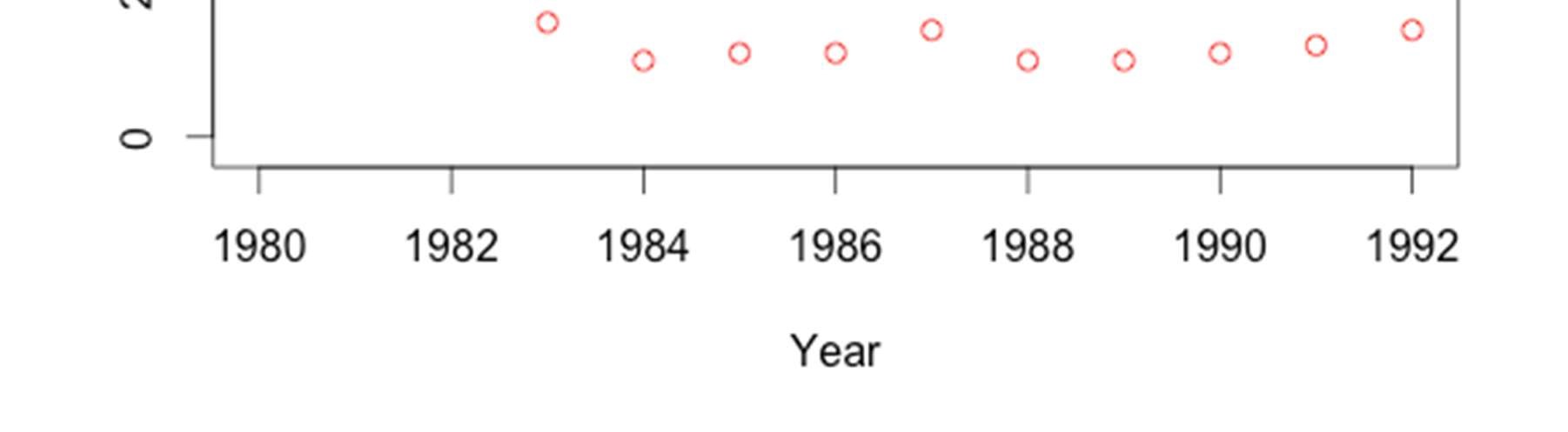
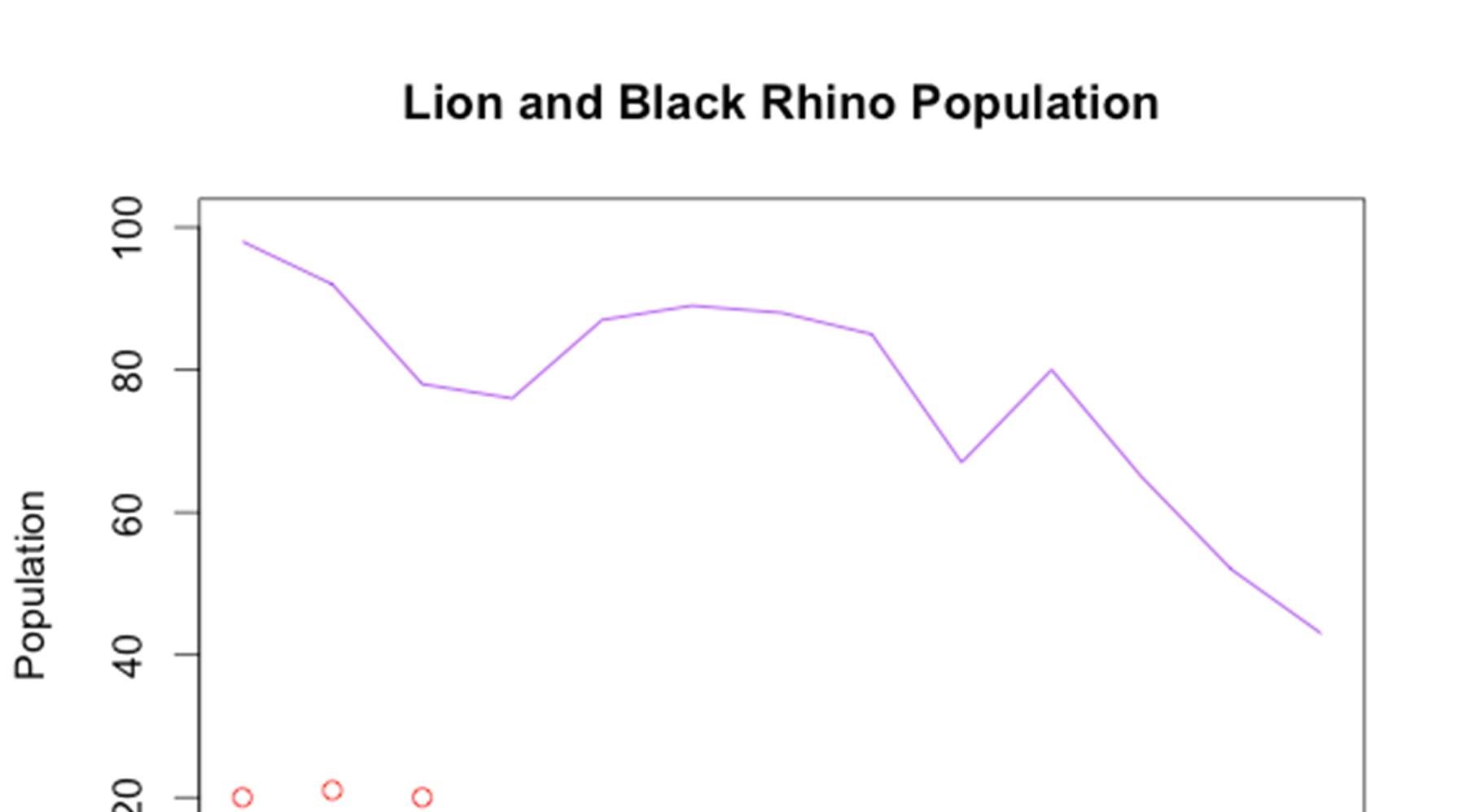
> Rhino2 <- subset(Rhino, Rhino$Year < 1993)

1. Create a line plot of the lion population sizes. On the same plot add the black rhino population sizes as points. Make the lion population line purple and the black rhino population points red. Includes a title and axis labels. You will need to set the limits of the y-axis from (0,100) otherwise the rhino points will not be visible when you add them to the plot.

> plot(Lion2$Year, Lion2$Population, type = "l", col = "purple", main = "Lion and Black

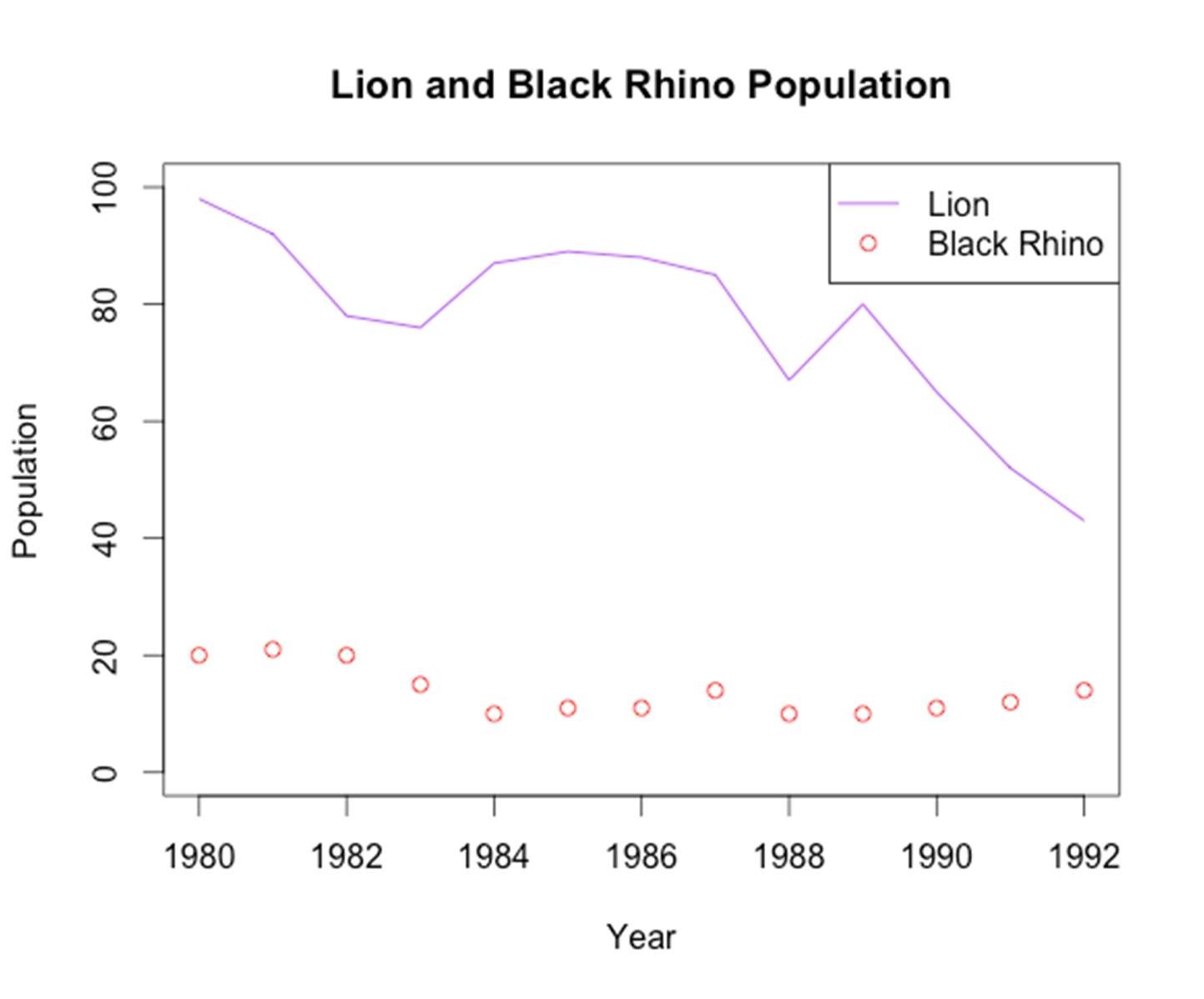
Rhino Population", xlab = "Year", ylab = "Population", ylim = c(0,100))

> points(Rhino2$Year, Rhino2$Population, col = "red")



1. Add a legend to the top right of the plot you just made – make sure next to lion you have a purple line, and next to black rhino you have a red point.

> legend("topright", legend = c("Lion", "Black Rhino"), col = c("purple", "red"), lty = c(1,NA), pch = c(NA,1))

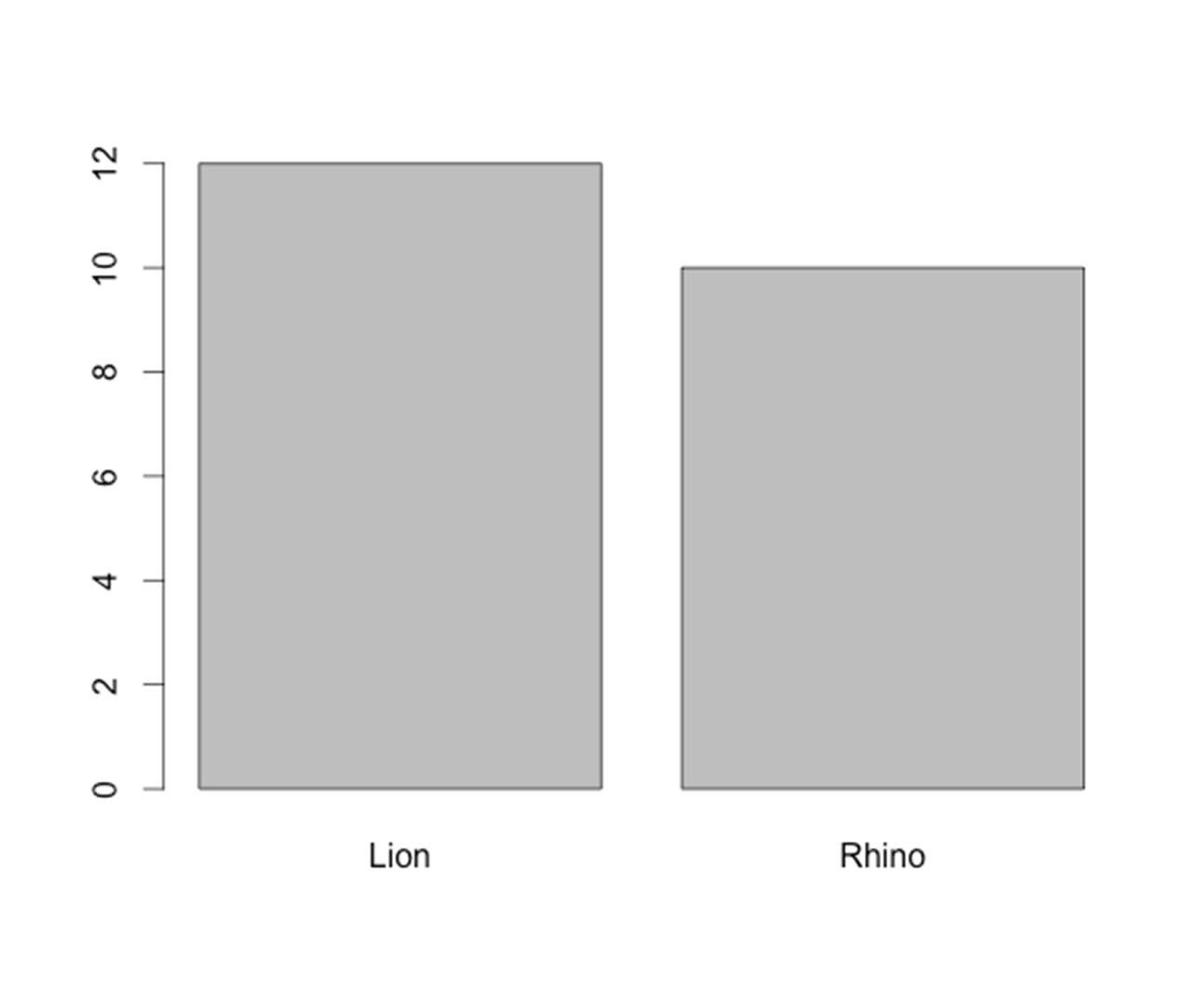


1. Create a barplot showing the minimum population size for black rhino and lion. Make sure to include a title and labels for each of the bars.

> data <- c(min(Lion$Population), min(Rhino$Population))

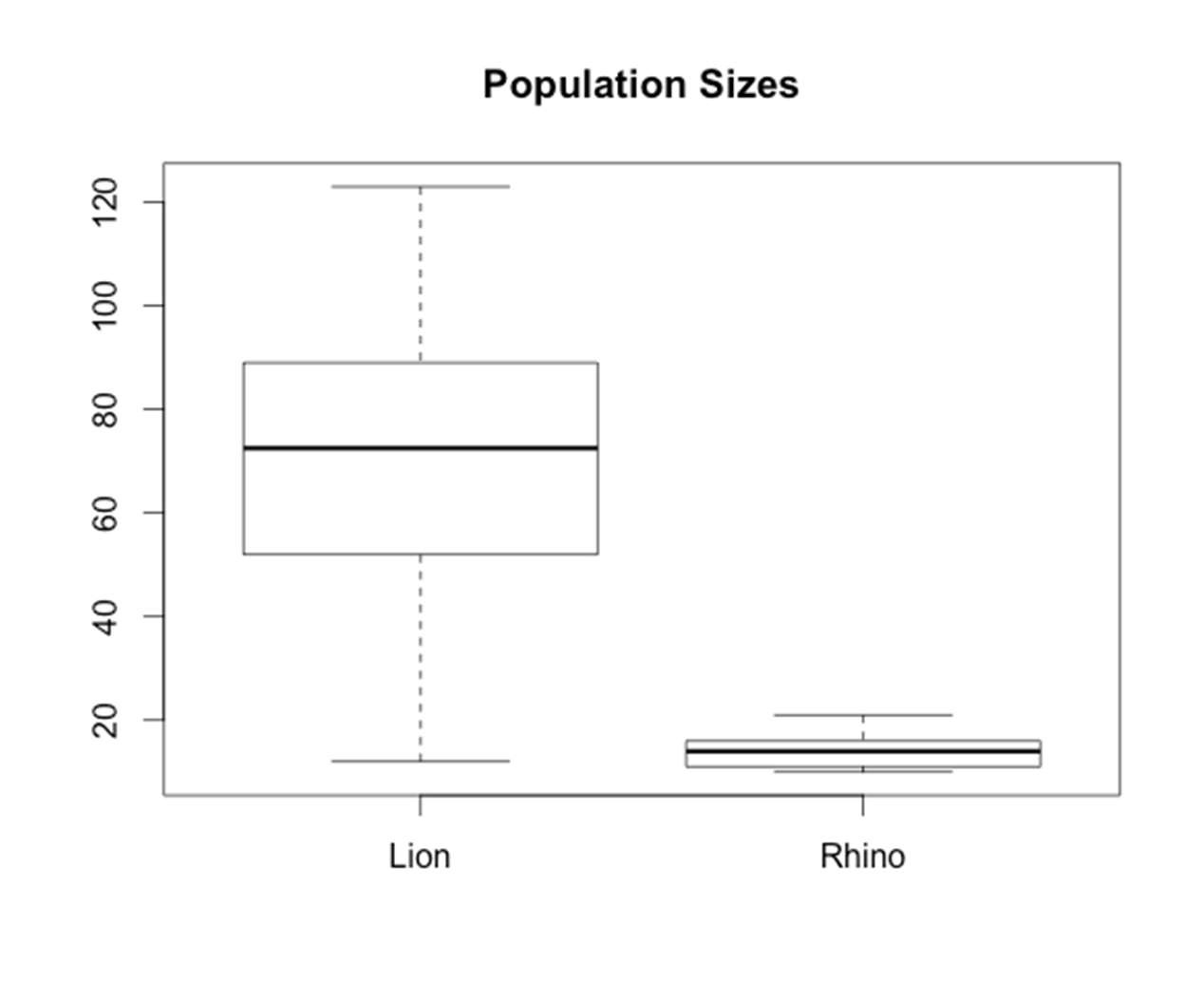
> names(data) <- c("Lion", "Rhino")

> barplot(data)



1. Create a side-by-side boxplot of the population size for lions and black rhinos – make sure to include a title, and a label below each box to show which animal it represents.

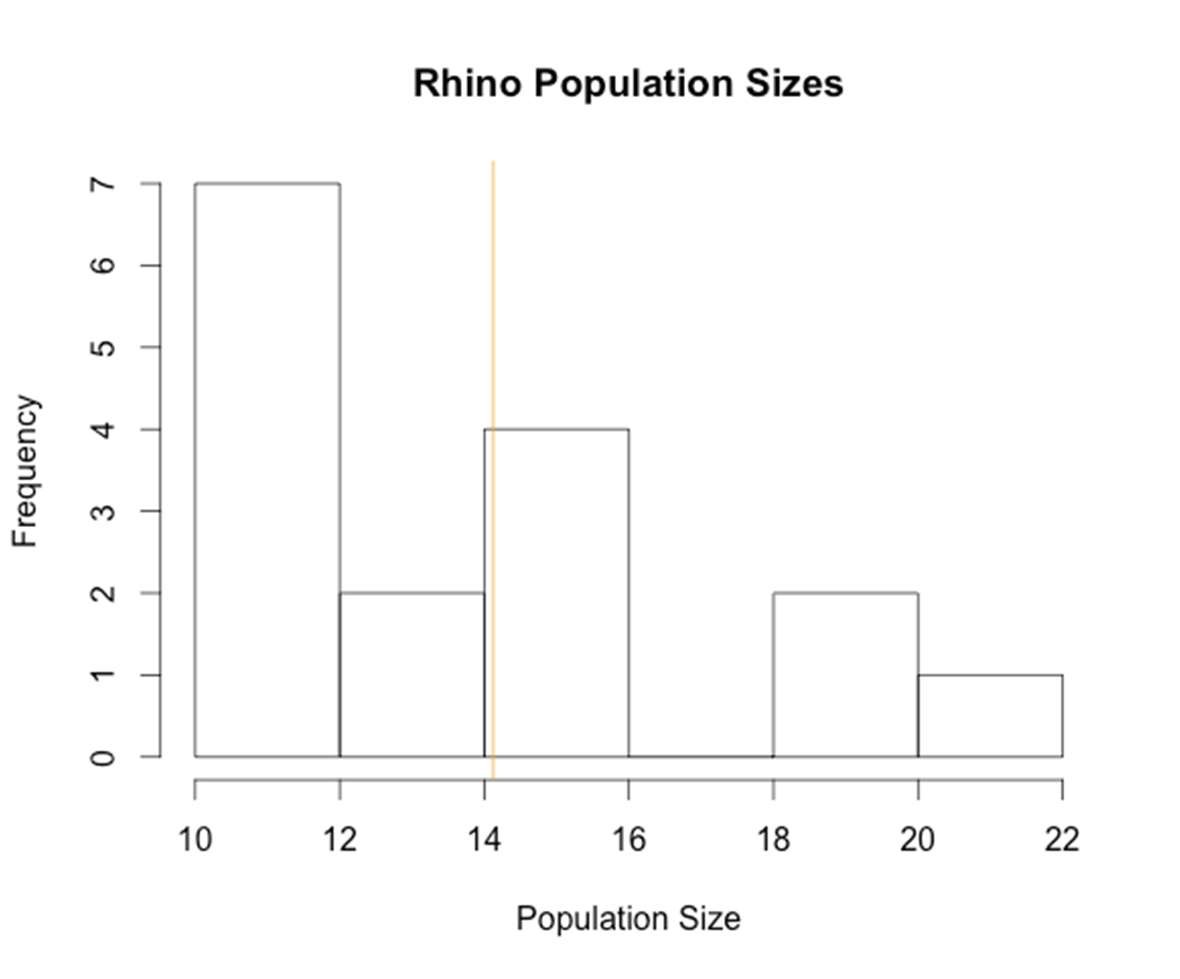
> boxplot(Lion$Population, Rhino$Population, names = c("Lion", "Rhino"), main = "Population Sizes")



1. Create a histogram of the population size for black rhinos. Add an orange vertical line to show the mean population size.

> hist(Rhino$Population, main = "Rhino Population Sizes", xlab = "Population Size")

> abline(v=mean(Rhino$Population), col = "orange")



Data Sources:

Brooks, B., Traill, L. W., and Bradshaw, C. J. A. 2006. Minimum viable population sizes and global extinction risk are unrelated. Ecology Letters 9 (4): 375-382. White Rhinocerous (Ceratotherium simum) population numbers from Matobo National Park, Zimbabwe from 1962-1994.

Lion (Panthera leo) population numbers from Ngorongoro Crater, Tanzania from 1963-1992.

Black Rhino (Diceros bicornis) population numbers from Ngorongoro Crater, Tranzania from 1980-1995.

Dublin, H. T., Sinclair, A. R. E., Boutin, S., Anderson, E., Jago, M., and Arcese, P. 1990. Does competition regulate ungulate populations? Further evidence from Serengeti, Tanzania. Oecologia 82: 283-288.

Blue Wildebeest (Connochaetes taurinus) population numbers in the Serengeti Plains from 1961-1986.

African Buffalo (Syncerus caffer) population numbers in the Serengeti Plains from 1965-1976.

Hanby, J. P., Boycott, J. D. and Packer, C. 1995. Ecology, demography, and behavior of lions in two contasting habitat: Ngorongoro Crater and the Serengeti Plains. In:

Serengeti II by Sinclair, A. R. E. and Arcese, P. pp 315-331.

Lion (Panthera leo) population numbers in the Serengeti Plains from 19661990.

Teleki, G., Hunt Jr., E. E. and Pfifferling, J. H. 1976. Demographic Observations (19631973) on the Chimpanzees of Gombe National Park, Tanzania.

Chimpanzee (Pan troglodytes) population numbers in Gombe National Park, Tanzania from 1964-1973.