5. Vectors, sequences, and logical operators

(a) Assign the names `x` and `y` to the values 5 and 7, respectively. Find x^y and assign the result to `z`. What is the valued stored in `z`?

(b) Create the vectors `u` $= (1, 2, 5, 4)$ and `v` $= (2, 2, 1, 1)$ using the `c()` function.

(c) Provide R code to find which component of `u` is equal to 5.

(d) Provide R code to give the components of `v` greater than or equal to 2.

(e) Find the product `u` $\times$ `v`. How does R perform the operation?

(f) Explain what R does when two vectors of unequal length are multiplied together. Specifically, what is `u` $\times$ `c(u, v)`?

(g) Provide R code to define a sequence from 1 to 10 called `G` and subsequently to select the first three components of `G`.

(h) Use R to define a sequence from 1 to 30 named `J` with an increment of 2 and subsequently to choose the first, third, and eighth values of `J`.

(i) Calculate the scalar product (dot product) of $q = (3, 0, 1, 6)$ by $r = (1, 0, 2, 4)$.

(j) Define the matrix $\mathbf{X}$ whose rows are the `u` and `v` vectors from part (b).

(k) Define the matrix $\mathbf{Y}$ whose columns are the `u` and `v` vectors from part (b).

(l) Find the matrix product of $\mathbf{X}$ by $\mathbf{Y}$ and name it $\mathbf{W}$.

(m) Provide R code that computes the inverse matrix of $\mathbf{W}$ and the transpose of that inverse.

Solution:

(a) The valued stored in `z` is 78125.

```
> x <- 5
> y <- 7
> z <- x^y
> z

[1] 78125
```

(b)

```
> u <- c(1, 2, 5, 4)
> v <- c(2, 2, 1, 1)
```

(c)

```
> which(u == 5)

[1] 3
```

(d)

```
> which(v >= 2)

[1] 1 2
```

(e) Multiplication of vectors with R is element by element.

```
> uv <- u * v
> uv

[1] 2 4 5 4
```

(f) The values in the shorter vector are recycled until the two vectors are the same size. In this case, `u*c(u, v)` is the same as `c(u, u)*c(u, v)`.

```
> u * (c(u, v))

[1]  1  4 25 16  2  4  5  4

> c(u, u) * c(u, v)

[1]  1  4 25 16  2  4  5  4
```

(g)

```
> G <- 1:10
> G[1:3]

[1] 1 2 3
```

(h)

```
> J <- seq(from = 1, to = 30, by = 2)
> J[c(1, 3, 8)]

[1]  1  5 15
```

(i)

```
> q <- c(3, 0, 1, 6)
> r <- c(1, 0, 2, 4)
> q %*% r

     [,1]
[1,]   29
```

(j)

```
> X <- rbind(u, v)
> X

  [,1] [,2] [,3] [,4]
u    1    2    5    4
v    2    2    1    1
```

(k)

```
> Y <- cbind(u, v)
> Y

     u v
[1,] 1 2
[2,] 2 2
[3,] 5 1
[4,] 4 1
```

(l)

```
> W <- X %*% Y
> W

   u  v
u 46 15
v 15 10
```

(m)

```
> solve(W)

            u           v
u  0.04255319 -0.06382979
v -0.06382979  0.19574468
```

```
> t(solve(W))

            u           v
u  0.04255319 -0.06382979
v -0.06382979  0.19574468
```

5. The data frame VIT2005 in the PASWR2 package contains descriptive information and the appraised total price (in euros) for apartments in Vitoria, Spain.

(a) Create a frequency table, a piechart, and a barplot showing the number of apartments grouped by the variable out. For you, which method conveys the information best?

(b) Characterize the distribution of the variable totalprice.

(c) Characterize the relationship between totalprice and area.

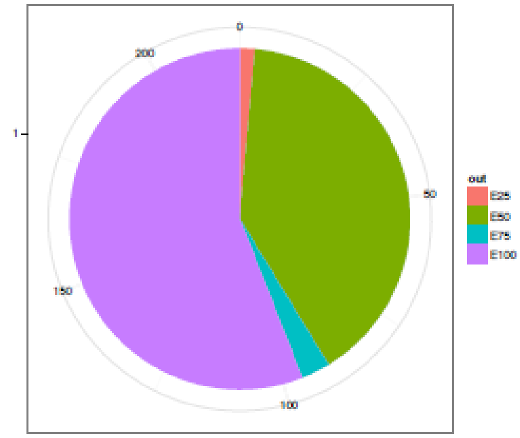
(d) Create a Trellis plot of totalprice versus area conditioning on toilets. Create the same graph with ggplot2 graphics. Are there any outliers? Ignoring any outliers, between what two values of area do apartments have both one and two bathrooms?

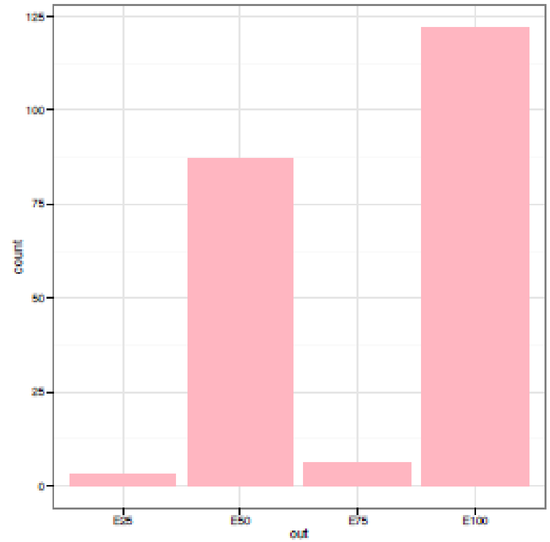
(e) Use the area values reported in (d) to create a subset of apartments that have both one and two bathrooms. By how much does an additional bathroom increase the appraised value of an apartment? Would you be willing to pay for an additional bathroom if you lived in Vitoria, Spain?

Solution:

(a) The barplot is the easiest to read.

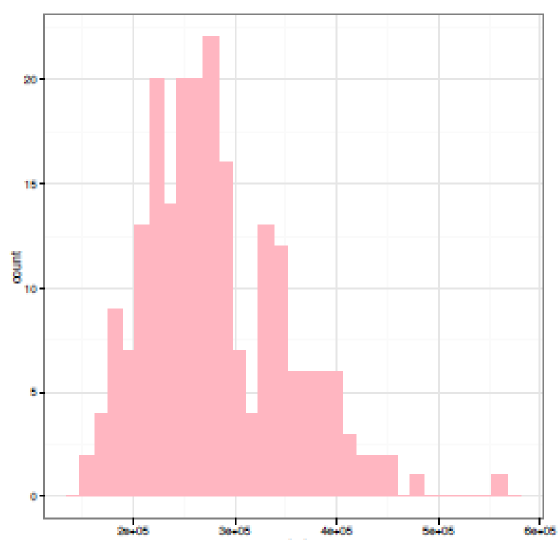
```
> VIT2005$out <- factor(VIT2005$out,
+                       levels = c("E25", "E50", "E75", "E100"))
> levels(VIT2005$out)

[1] "E25"  "E50"  "E75"  "E100"

> xtabs(~out, data = VIT2005)

out
 E25  E50  E75 E100
   3   87    6  122

> p1 <- ggplot(data = VIT2005, aes(x = factor(1), fill = out)) +
+   geom_bar(width = 1)
> p1 + coord_polar(theta = "y") +
+   theme_bw() +
+   labs(x = "", y = "")
> p2 <- ggplot(data = VIT2005, aes(x = out)) +
+   geom_bar(fill = "lightpink") +
+   theme_bw()
> p2
```





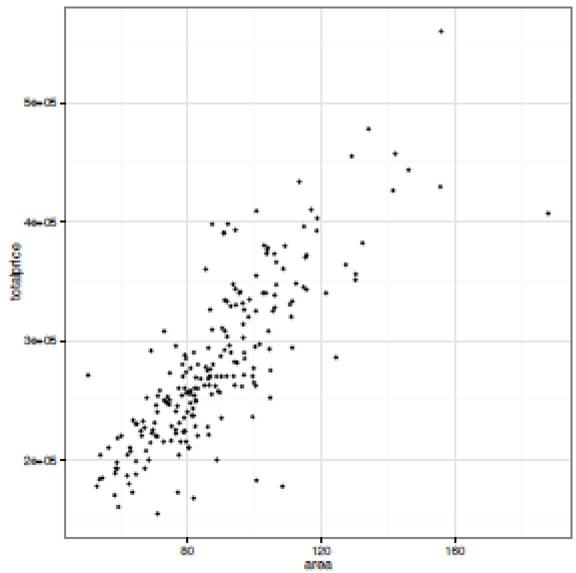
(b) The distribution of `totalprice` is skewed to the right with one outlier (€560000). The median `totalprice` is €269750 and the IQR for `totalprice` is €100125.

```
> ggplot(data = VIT2005, aes(x = totalprice)) +
+   geom_histogram(fill = "lightpink") +
+   theme_bw()
> max(VIT2005$totalprice)

[1] 560000

> median(VIT2005$totalprice)

[1] 269750

> IQR(VIT2005$totalprice)

[1] 100125
```



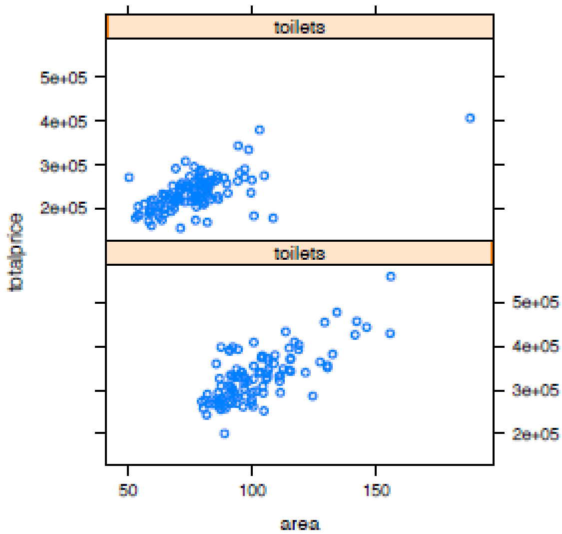
(c) There is a positive linear relationship between `totalprice` and `area`.

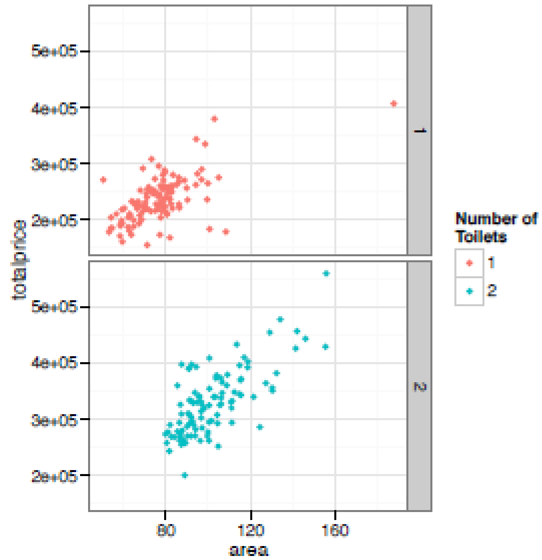
```
> ggplot(data = VIT2005, aes(x = area, y = totalprice)) +
+   geom_point() +
+   theme_bw()
```



(d) Apartments with one bathroom are generally between 50 and 100 m^2, while apartments with two bathrooms are generally between 80 and 160 m^2. The intersection of apartments with one and two bathrooms is roughly 80 to 100 m^2.

```
> xyplot(totalprice ~ area | toilets, data = VIT2005, layout = c(1, 2),
+        as.table = TRUE)
> TEXT <- "Number of\nToilets"
> ggplot(data = VIT2005, aes(x = area, y = totalprice,
+                            color = as.factor(toilets))) +
+   geom_point() +
+   facet_grid(toilets ~ .) +
+   theme_bw() +
+   guides(color = guide_legend(TEXT))
```





(e) The median increase in `totalprice` for a second bathroom for apartments between 80 and 100 m^2 is €36000. Answers will vary for answering whether readers would be willing to spend €36000 for an additional bathroom.

```
> bothbaths <- subset(VIT2005, subset = area >= 80 & area <= 100)
> ANS <- tapply(bothbaths$totalprice, bothbaths$toilets, median)
> ANS

     1      2
255000 291000

> diff(ANS)

    2
36000
```