

## DATA 202/472 – MODULE 2 EXERCISES

**Part I.** A company has a database that records its sales. The database is comprised of two tables. Table `staff` lists staff and their locations. Table `sales` records sales made. The tables contain the following data.

Table `staff`:

staff_id	name	location
101	James	Auckland
102	William	Wellington
103	Thomas	Christchurch
104	Olivia	Wellington
105	Jack	Auckland

Table `sales`:

invoice	staff_id	description	price
9753	101	Honda Fit Hybrid	14900
9754	102	Suzuki Swift 12XG	13500
9755	103	Suzuki Swift HYBRID RS	19999
9756	101	Mazda Demio 13-Skyactive	11000
9757	104	Nissan March S	9900
9758	105	Toyota Vitz 5D F	8950
9759	102	Toyota Prius S	24000
9760	101	Suzuki Swift 12XG	20000

a. Write down the output of the following SQL query:

```
SELECT name, description, price
FROM staff
INNER JOIN sales ON sales.staff_id=staff.staff_id
WHERE location = 'Wellington'
```

(3 Marks)

b. Write down the output of the following SQL query:

```
SELECT staff.staff_id, name, count(*) AS total_sales
FROM staff
INNER JOIN sales ON sales.staff_id=staff.staff_id
GROUP BY staff.staff_id
ORDER BY total_sales
```

(3 Marks)

c. Write a SQL query that returns the following table:

location	sum_sales
Auckland	54850
Christchurch	19999
Wellington	47400

(3 Marks)

d. Write a SQL query that returns **only** the description of the item with the highest price:

(3 Marks)

---

**Part II.** For the following questions, assume that variable `rents` contains a data frame about weekly market rent for properties in some areas of Wellington. The whole content of the data frame is shown in the table below.

Table `rents`:

area	bedrooms	lower_quartile	median_rent	upper_quartile
Aro Valley	1	378	400	458
Aro Valley	2	405	500	563
Aro Valley	3	660	695	790
Karori	1	375	415	440
Karori	2	495	560	580
Northland	1	388	420	448
Northland	2	500	510	545
Northland	3	645	675	764
Island Bay	1	388	400	448
Island Bay	2	515	560	600
Seatoun	1	430	493	535

- a. Write R code that adds a column named `IQR` to the `rents` data frame which records the difference in `upper_quartile` and `lower_quartile` (hint: using `dplyr`):

(3 Marks)

- b. Write R code to display the row(s) in which `IQR` is maximum:

(3 Marks)

- c. Write R code using the pipe operator `%>%` to change the name of column bedrooms to size, and then display the market rent information in Karori and Northland.

(3 Marks)

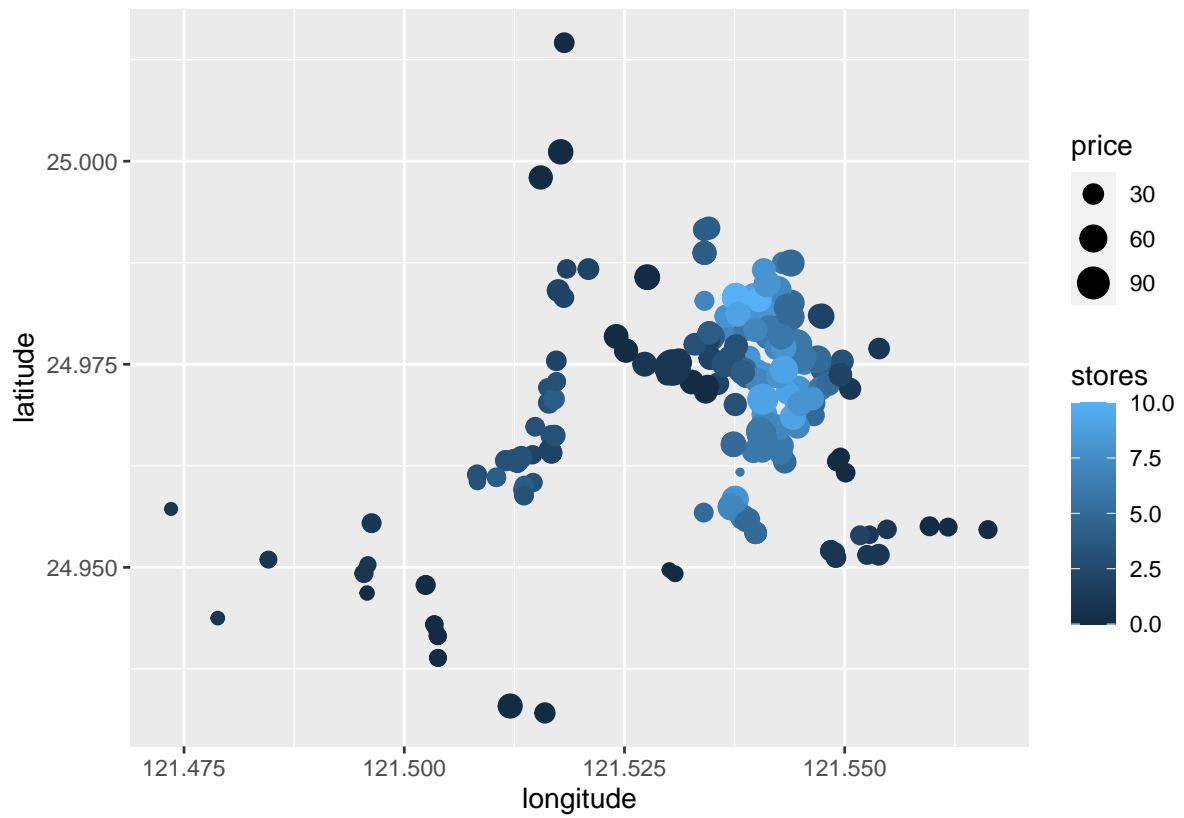
- d. Write down the output of the following code:

```
rents[rents$median_rent > 600, 1]
```

(3 Marks)

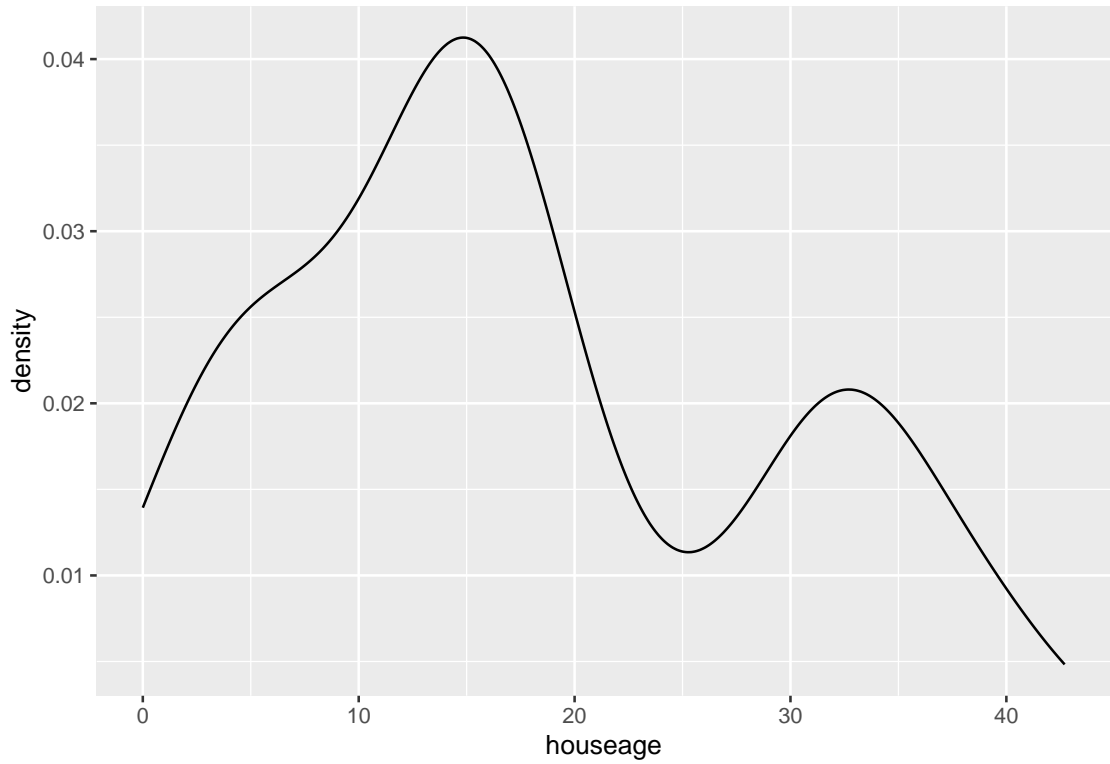
**Part III.** You are given a dataset called housing, part of which is shown here:

##	houseage	distMRT	stores	latitude	longitude	price
## 1	0.0	292.99780	6	24.97744	121.5446	69.7
## 2	19.1	461.10160	5	24.95425	121.5399	34.0
## 3	6.4	90.45606	9	24.97433	121.5431	62.2
## 4	4.5	2275.87700	3	24.96314	121.5115	29.3
## 5	35.3	614.13940	7	24.97913	121.5367	33.1



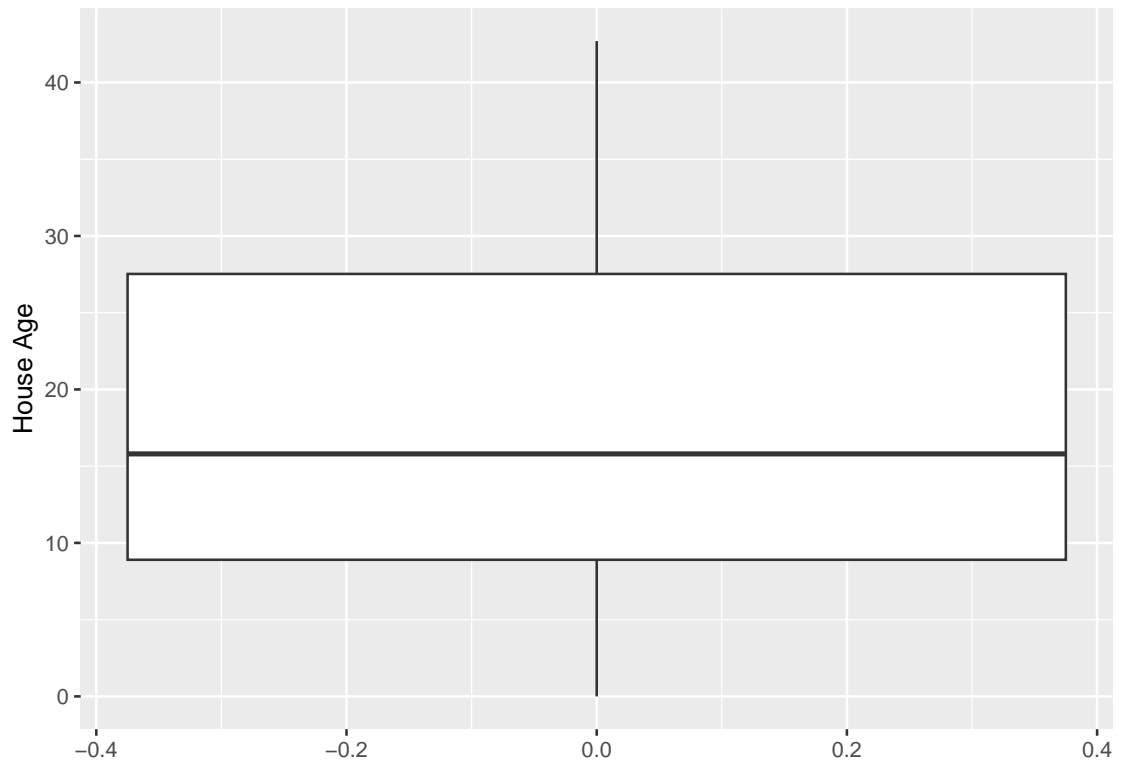
a. Examine the plot above and write R code to produce that plot:

(4 Marks)



b. Examine the plot above and write R code to produce that plot:

(3 Marks)



c. Examine the plot above and write R code to produce that plot:

(3 Marks)