MA20277 Coursework 1

23011

```r
knitr::opts_chunk$set(echo = TRUE)
```

```r
library(dplyr, warn.conflicts = FALSE)
library(lubridate, warn.conflicts = FALSE)
library(tidyr)
library(ggplot2)
library(patchwork)
```

Question 1 [19 marks]

An orchid grower delivered a large sample of orchids to a distributor on 20 October 2022. Each orchid's height was recorded in inches and each orchid was assigned a score between 0 and 10 (0=very poor quality, 10=excellent quality). Any orchid with a score above 6 is bought by the distributor, while a score of 6 or lower leads to the orchid not being bought by the distributor.

The orchid grower asks you to analyze the data they collected. In addition to the height and score, you are given the type of orchid, the temperature at which the plant was grown, the levels of phosphate, potassium and sulfur levels used for fertilization, and the date the orchid was transferred to an individual pot in spring.

The full data are in the file "Orchids.csv" and a detailed data description is provided in the file "Data Descriptions.pdf".

a) *Load and clean the data. Extract and provide the first two rows of the data set. State the minimum and maximum observed phosphate, potassium and sulphur levels.* [**4 marks**]

Loading in `Orchids.csv`

```r
Orchids=read.csv("Orchids.csv")
```

Cleaning Data Set:

```r
glimpse(Orchids)
```

```
## Rows: 14,917
## Columns: 8
## $ Height   <dbl> 16.3, 2.6, 22.7, 6.7, 6.4, 6.3, 9.9, 7.2, 8.9, 14.3, 16.2, 3.~
## $ Phos     <int> 89, 0, 83, 63, 99, 68, 82, 90, 77, 96, 111, 87, 83, 77, 90, 8~
## $ Potas    <int> 270, 265, 295, 353, 309, 272, 303, 242, 0, 272, 295, 291, 286~
## $ Sulf     <int> 38, 39, 34, 35, 43, 36, 38, 41, 35, 0, 35, 33, 34, 39, 37, 0,~
## $ Planting <chr> "2022-03-19", "2022-04-01", "2022-04-14", "2022-03-23", "2022~
## $ Type     <chr> "Phalaenopsis", "Odontoglossum", "Dendrobium", "Oncidium", "C~
## $ Temp     <dbl> 27.7, 18.1, 17.5, 22.6, 21.1, 21.7, 17.6, 21.6, 24.3, 17.1, 1~
## $ Quality  <int> 7, 5, 6, 7, 6, 3, 5, 6, 6, 4, 6, 7, 5, 7, 6, 7, 5, 5, 1, 7, 3~
```

Cleaning Data Set:

- `Planting`: Incorrect format (character not a date) and uninformative name.
- `Phos`, `Potas` and `Sulf`: Replace zeros with `NA`, since zero was used as missing value indicator

```r
Orchids$Planting = as_date( Orchids$Planting, format = "%Y-%m-%d")
Orchids = rename(Orchids, Date_Planted = Planting)
Orchids$Phos[Orchids$Phos == 0] <- NA
Orchids$Potas[Orchids$Potas == 0] <- NA
Orchids$Sulf[Orchids$Sulf == 0] <- NA
```

First two rows of the data set `Orchids` cleaned.

```r
Orchids %>% slice_head(n=2)
```

```
##   Height Phos Potas Sulf Date_Planted          Type Temp Quality
## 1   16.3   89   270   38   2022-03-19  Phalaenopsis 27.7       7
## 2    2.6   NA   265   39   2022-04-01 Odontoglossum 18.1       5
```

The minimum and maximum observed Phosphate, Potassium and Sulfur levels.

```r
#Creating Orchids_tidy data frame for later usage
Orchids_tidy = Orchids %>%
  pivot_longer( cols = Phos:Sulf, names_to='Fertilizer') %>%
  rename('Level' = value)
Orchids_tidy %>% filter(`Level`>0) %>%
  group_by(Fertilizer) %>%
  summarize('Maximum Level' = max(`Level`), 'Minimum Level' = min(`Level`))
```

```
## # A tibble: 3 x 3
##   Fertilizer `Maximum Level` `Minimum Level`
##   <chr>                <int>           <int>
## 1 Phos                   130              46
## 2 Potas                  385             195
## 3 Sulf                    46              28
```

The maximum/minimum phosphate levels are 130/46. The maximum/minimum potassium levels are 385/195. The maximum/minimum sulfur levels are 46/28.

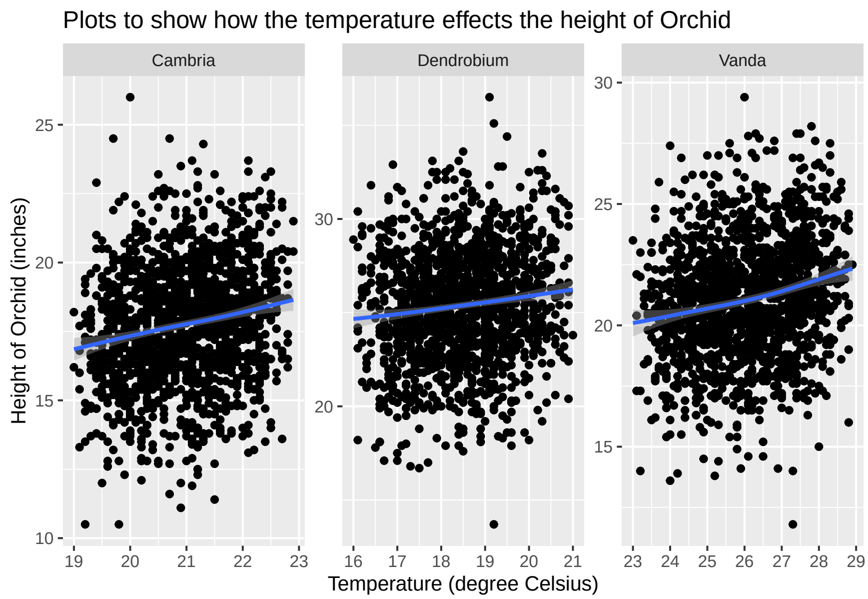
b) *Explore the relationship of temperature and plant height for the three types of orchid with the highest average height. Further investigate how these three types compare regarding their quality.* [**5 marks**]

```r
Orchids %>% group_by(Type) %>%
  summarise( 'Average Height' = mean(Height)) %>%
  arrange(desc(`Average Height`)) %>% slice(1:3)
```

```
## # A tibble: 3 x 2
##   Type       `Average Height`
##   <chr>                 <dbl>
## 1 Dendrobium             25.4
## 2 Vanda                  21.1
## 3 Cambria                17.8
```

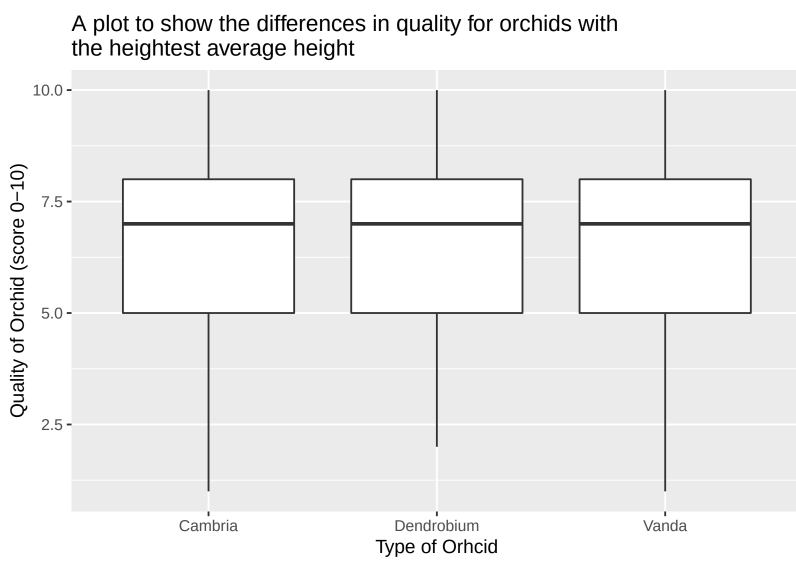
The 3 orchids with the highest average height are `Dendrobium`, `Vanda` and `Cambria`

```r
Orchids %>% filter(Type %in% c('Dendrobium', 'Vanda', 'Cambria' )) %>%
  ggplot(aes(x = Temp, y=Height)) + facet_wrap(~Type, scales = 'free') +
  geom_point() + geom_smooth() +
  labs(x = 'Temperature (degree Celsius)', y = 'Height of Orchid (inches)',
       title= 'Plots to show how the temperature effects the height of Orchid')
```



All three plots show variation for the heights and temperatures of various orchids. All three Orchids are also grown at different temperatures. `Cambria` ranges from 19-23 Celsius, `Dendrobium` ranges from 16-21 Celsius and `Vanda` ranges from 23-29 Celsius. There is a clear positive correlation between temperature and height (as temperature increases, so does the height of the orchid).

```
Orchids %>% filter(Type %in% c('Dendrobium', 'Vanda', 'Cambria' )) %>%
  ggplot(aes(x = Type, y=Quality)) + geom_boxplot() +
  labs(x = 'Type of Orhcid', y = 'Quality of Orchid (score 0-10)', title=
         'A plot to show the differences in quality for orchids with
the heightest average height')
```



The interquartile range and median quality are approximately the same for each type of orchid. This implies that there are no significant differences regarding the quality of orchid produced. Interestingly, `Dendrobium` has a higher minimum quality than `Cambria` and `Vanda`.

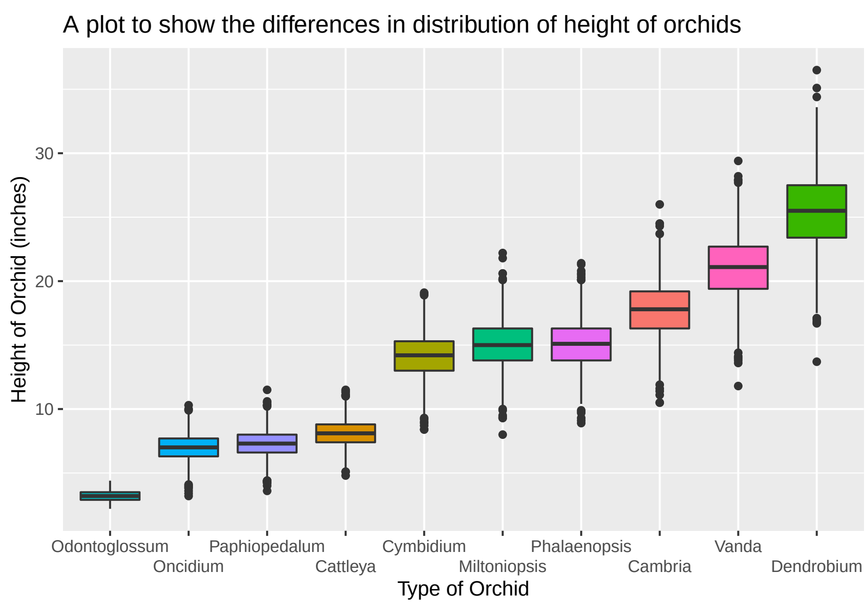
c) *Investigate differences between the types of orchids in terms of their distribution of height. Are there any differences in growing conditions?* [**5 marks**]

The growing conditions I will consider are fertilizers and temperature.

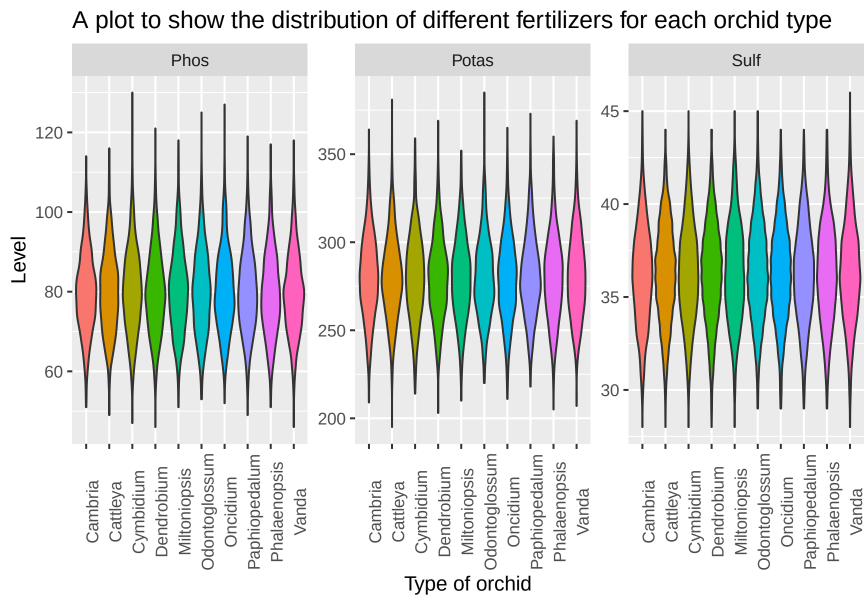
```r
Orchids_tidy %>% group_by(Type) %>% summarize('Average Height (inches)' = mean(Height),
            'Average Temperature (degree Celsius)' = mean(Temp)) %>%
  arrange(desc(`Average Temperature (degree Celsius)`))
```

```
## # A tibble: 10 x 3
##    Type          `Average Height (inches)` `Average Temperature (degree Celsiu~`
##    <chr>                             <dbl>                                 <dbl>
##  1 Vanda                             21.1                                   26.1
##  2 Phalaenopsis                      15.1                                   26.0
##  3 Oncidium                           6.99                                  21.0
##  4 Paphiopedalum                      7.31                                  21.0
##  5 Cattleya                           8.09                                  21.0
##  6 Cambria                           17.8                                   21.0
##  7 Cymbidium                         14.2                                   18.5
##  8 Miltoniopsis                      15.0                                   18.5
##  9 Odontoglossum                      3.20                                  18.5
## 10 Dendrobium                        25.4                                   18.5
```

```r
Orchids %>% ggplot(aes(x=reorder(Type,Height), y= Height, fill = Type)) +
  geom_boxplot() + guides(x = guide_axis(n.dodge=2)) +
  labs(x = 'Type of Orchid', y = 'Height of Orchid (inches)', title=
       'A plot to show the differences in distribution of height of orchids') +
  theme(legend.position = "")
```




```r
Orchids_tidy %>% ggplot(aes(x =Type , y = Level, fill = Type)) +
  facet_wrap(~Fertilizer, scales = 'free') + geom_violin() +
  labs(x = 'Type of orchid', y = 'Level', title=
     'A plot to show the distribution of different fertilizers for each orchid type') +
  theme(axis.text.x = element_text(angle =90), legend.position = "")
```



The box plots above indicate that there is a fair distribution in heights of orchids. Considering orchid types in ascending average height, as the average height increases, so does the interquartile range, and the range between the maximum and minimum values of height. There are two groups of orchids which all have a very similar interquartile range and average height. From the data frame created above, observe that the groups of orchids also have differences and similarities in temperature.

- `Cymbidium`, `Miltoniopsis` and `Phalaenopsis`: Interestingly there are differences in growing conditions; `Cymbidium` and `Miltoniopsis` have an approximate average temperature of 18.5 degrees, whereas `Phalaenopsis` has a significantly higher average temperature of 26.0 degrees.
- `Cattleya`, `Oncidium` and `Paphiopedalum`; they all have approximately the same average temperature at 21.0 degrees.
- The Orchid with the lowest average height is `Odontoglossum`, and the three orchids with the highest average height, as mentioned before, are `Dendrobium`, `Vanda` and `Cambria`. Interestingly, these three orchid types have differences in growing conditions. There are varying average temperatures from 18.5, 26.1 and 21.0 degrees respectively.

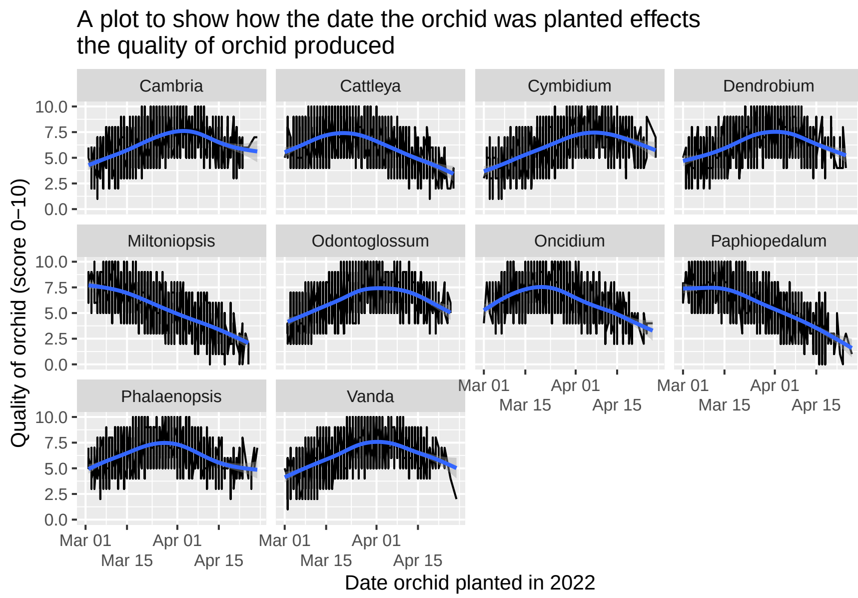
Observe from the violin plots above that the distribution of the different fertilizers the same across the orchid types. The minimum and maximum levels of fertilizers used for different types of orchid are also very similar. There are no differences in fertilizer levels used when growing each type of orchid. Hence, there are differences in temperature at which each orchid type is grown at, but there are no differences in the fertilizer levels used to grow each orchid type.

d) *The orchid grower wants to optimize the times at which the different types of orchids are transferred to individual pots. The aim is to have a large proportion of orchids being bought by the distributor. Use the data to advise the orchid grower on which two types of orchids they should plant first in 2023. When should the first orchid be planted? Discuss which assumption you make when basing your suggestions on the data.* [**5 marks**]

Choose to create line plots for each type of orchid. These are easier to interpret how the quality of the orchid produced was impacted from one day to the next day when each new orchid is planted. This would be harder to interpret on scatter plots.

```
ggplot(Orchids , aes(x= Date_Planted, y = Quality)) + geom_line() +
  facet_wrap(~Type) + geom_smooth() +
  guides(x = guide_axis(n.dodge=2)) +
  labs(x = 'Date orchid planted in 2022', y = 'Quality of orchid (score 0-10)', title=
```

```
        'A plot to show how the date the orchid was planted effects
the quality of orchid produced')
```



The LOESS curves suggest that the orchid grower should plant orchids `Miltoniopsis` and `Paphiopedalum` first in 2023. For all other orchids the LOESS curve gradient is positive at the start of the planting period. for `Miltoniopsis`, the gradient of the LOESS curve is decreasing from `Mar 01`. This implies that for every day from `Mar 01` the orchid grower delays planting `Miltoniopsis`, the quality of `Miltoniopsis` decreases. The gradient of the LOESS curve for `Paphiopedalum` is neither decreasing or increasing between `Mar 01` and `Mar 15`. This is important since it implies that any `Paphiopedalum` orchid planted from `Mar 01` to `Mar 15` will have the same maximum quality. This implies that the orchid grower should plant `Miltoniopsis` first, and should plant the first `Miltoniopsis` orchid on `Mar 01`. The main assumption made is that the growing conditions are the same as the previous year. It is important that the levels of fertilizer used and temperature stay the same throughout the growing time of the orchid, since we do not have data on how different levels of fertilizer and ranges of temperature effect the quality of types of orchids.

Question 2 [27 marks]

The country *Utopia* has collected data on their ambulance service and the patients admitted to the country's hospitals. The health department of Utopia has given you access to their data in the files "Ambulance.csv" and "Hospital.csv", and a data description is provided in the file "Data Descriptions.pdf". You are asked to consider the following tasks which are aimed towards analyzing the performance of their ambulance service and the factors influencing health outcomes:

a) *At which time of the day do we tend to see the highest frequency of calls to the ambulance service? Which proportion of calls leads to the patient being delivered to hospital?* **[4 marks]**

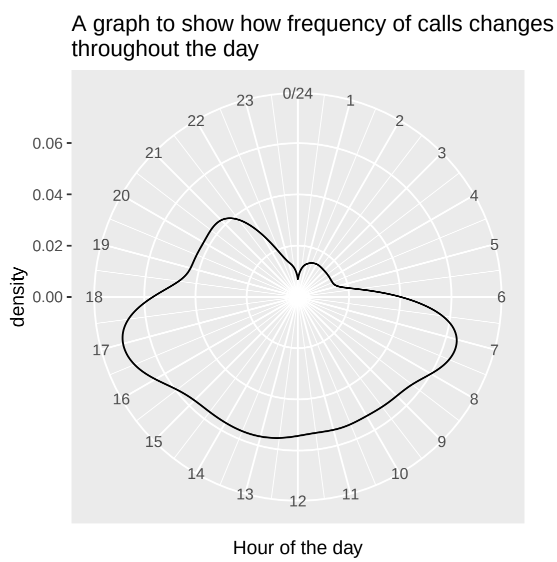
Loading in data

```
Ambulance=read.csv("Ambulance.csv")
Hospital=read.csv("Hospital.csv")
```

- `Call`, `Arrival` and `Hospital` : Incorrect format (should be date/time format) and uninformative names
- `Category1` and `Category2`: Incorrect format (Should be factor/categorical format) and uninformative names

```r
Ambulance$Call = ymd_hms(Ambulance$Call)
Ambulance$Arrival = ymd_hms(Ambulance$Arrival)
Ambulance$Hospital = ymd_hms(Ambulance$Hospital)
Ambulance$Category1 = as.factor(Ambulance$Category1)
Ambulance$Category2 = as.factor(Ambulance$Category2)
Ambulance = rename(Ambulance, Ambulance_Called = Call, Ambulance_Arrived = Arrival,
  Arrived_Hospital=Hospital, Initial_Category = Category1, Arrival_Category = Category2)
```

```r
#Converting Time from %H:%M:%S to numerical format
Ambulance$Time_Numeric=hour(Ambulance$Ambulance_Called) +
  (minute(Ambulance$Ambulance_Called)/60) +
  (second(Ambulance$Ambulance_Called)/60^2)
#plotting frequency density polar plot with hours of the day on perimeter.
ggplot(Ambulance, aes(x = Time_Numeric))+ coord_polar(theta = 'x') +
  geom_density() + labs( x = 'Hour of the day',
  title = 'A graph to show how frequency of calls changes
throughout the day') +
  scale_x_continuous(breaks=seq(0,24,1), limits=c(0,24) )
```



The region of time of day with the highest frequency of calls is 16:00 to 18:00.

```r
#calculating proportion of calls that did not lead to hospital admission,
#correlates to all NA values in Arrived_Hospital column
1-sum(is.na(Ambulance$Arrived_Hospital))/67172
```

```
## [1] 0.8002888
```

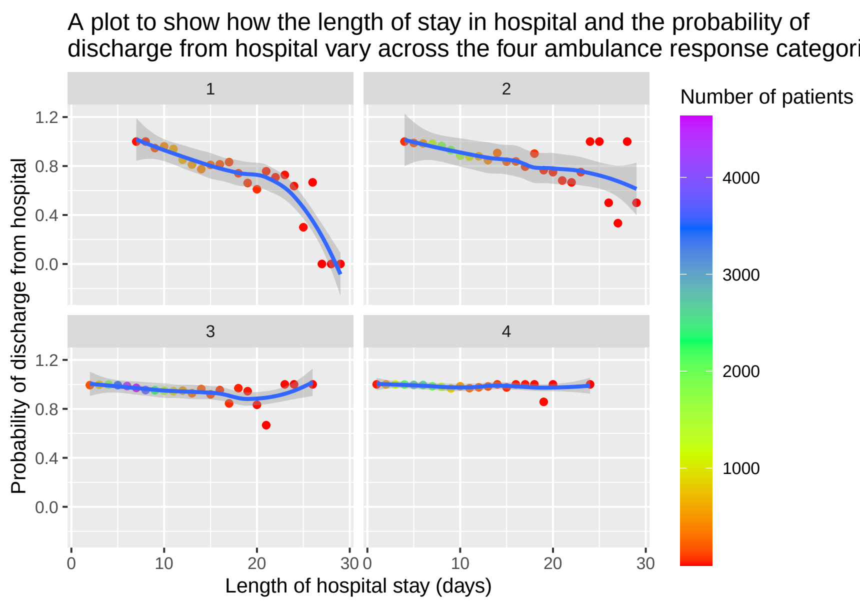
```r
#one minus proportion above since aim is calculating the complement proportion.
```

80% of calls to the ambulance service lead to the patient being delivered to hospital.

b) *How does the length of stay in hospital and the probability of discharge from hospital vary across the four ambulance response categories? Here, ambulance response category refers to that at the time of arrival of the ambulance.* [**4 marks**]

```r
Utopia = full_join( Ambulance, Hospital, by ="PatientID")
Utopia %>% drop_na(Arrived_Hospital) %>%
  group_by(Length, Arrival_Category) %>%
  summarise('Probability of discharge' = 1-mean(Outcome),
```

```
            'Number of patients' = n()) %>%
  ggplot(aes(x = Length, y= `Probability of discharge`,
             color = `Number of patients`, )) + geom_point()  +
  facet_wrap(~Arrival_Category) + geom_smooth() +
  theme(legend.key.height= unit(1.7, 'cm')) +
  scale_color_gradientn(colours = rainbow(5)) +
  labs(x = 'Length of hospital stay (days)', y =
         'Probability of discharge from hospital', title=
         'A plot to show how the length of stay in hospital and the probability of
discharge from hospital vary across the four ambulance response categories')
```



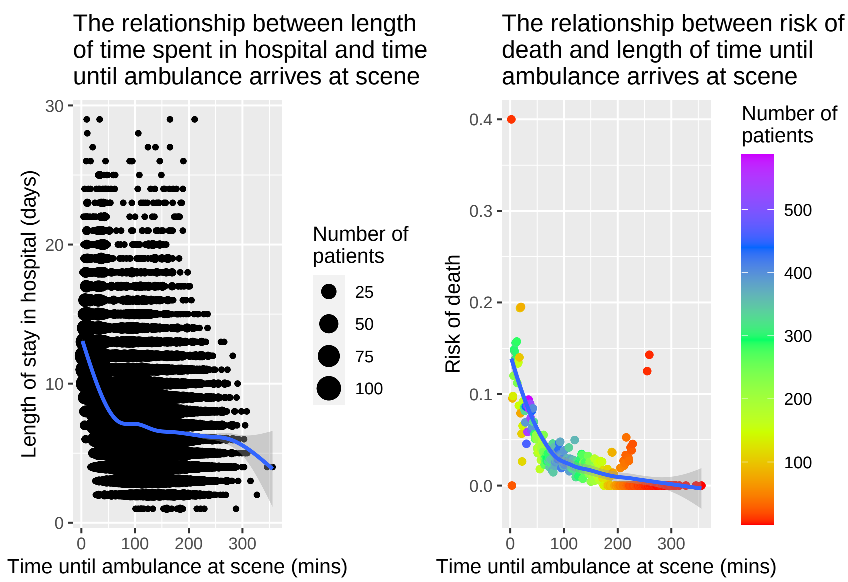
The probability of discharge and length of stay does vary between the four ambulance response categories.

- For ambulance response category 1, as the length of time in hospital increases, the probability of discharge dramatically decreases. Most of the patients in this category stay in hospital for 10 to 15 days. It is expected that patients who experience life threatening incidences tend to spend longer in hospital than patients from any other ambulance response category.
- For ambulance response category 2, as the length of time in hospital increases, the probability of discharge decreases significantly less than ambulance response category 1. Most of the patients in this category stay in hospital for 6 to 10 days.
- For ambulance response category 3 and 4, as the length of time in hospital increases, the probability of discharge stays at approximately the same level. Most of the patients in these categories stay in hospital for 2 to 10 days. It is expected that these patients tend to spend the least time in hospital considering that their incidences are less urgent than patients who have a lower ambulance response category.

c) *Does the data suggest that the length of stay in hospital and the risk of death increase with the time until the ambulance arrives, i.e, the length of time between calling the ambulance service and the ambulance arriving?* [**5 marks**]

```
#Ordering time differences in logical order
Utopia$`Time difference` = Utopia$Ambulance_Arrived -
  Utopia$Ambulance_Called
plot2 = Utopia %>% drop_na(Arrived_Hospital) %>%
  group_by(`Time difference`) %>%
  summarise('Risk of death' = mean(Outcome),'Number of patients' = n()) %>%
```

```
  ggplot(aes(x = `Time difference`, y = `Risk of death`, colour = `Number of patients`)) +
  geom_point() + geom_smooth() +
  theme(legend.key.height= unit(1.4, 'cm')) +
  scale_color_gradientn(colours = rainbow(5)) +
  labs(x = 'Time until ambulance at scene (mins)', title=
         'The relationship between risk of
death and length of time until
ambulance arrives at scene',
       y = 'Risk of death', colour = 'Number of \npatients' )
#editing colour and size of gradient legend on right hand side of plot.
plot1 = Utopia %>% drop_na(Arrived_Hospital) %>%
  ggplot(aes(y = Length, x = `Time difference`)) + geom_count() +
  geom_smooth() + labs(x = 'Time until ambulance at scene (mins)',
  title='The relationship between length
of time spent in hospital and time
until ambulance arrives at scene', y = 'Length of stay in hospital (days)') +
  scale_size_continuous(name = 'Number of \npatients')
plot1 + plot2
```



Key observations from graphs above:

- As the time taken for the ambulance to arrive at the scene increases, the risk of death decreases and the length of stay in hospital for the patient decreases.
- The ambulance service has a quicker response time to people with a higher risk of death. This suggests that Utopia's ambulance services is performing well at identifying people with a high risk of death from the initial call.

We can conclude that the patients with the highest risk of death spent longer in hospital, and these patients experienced a shorter delay in ambulance response time. This implies that the data suggests that the length of stay in hospital and the risk of death decrease as the time until the ambulance arrives at the scene increases.

d) *Make up your own question and answer it. Your question should be aimed towards understanding the factors influencing length of stay in hospital / health outcome. Originality will be rewarded.* [**7 marks**]

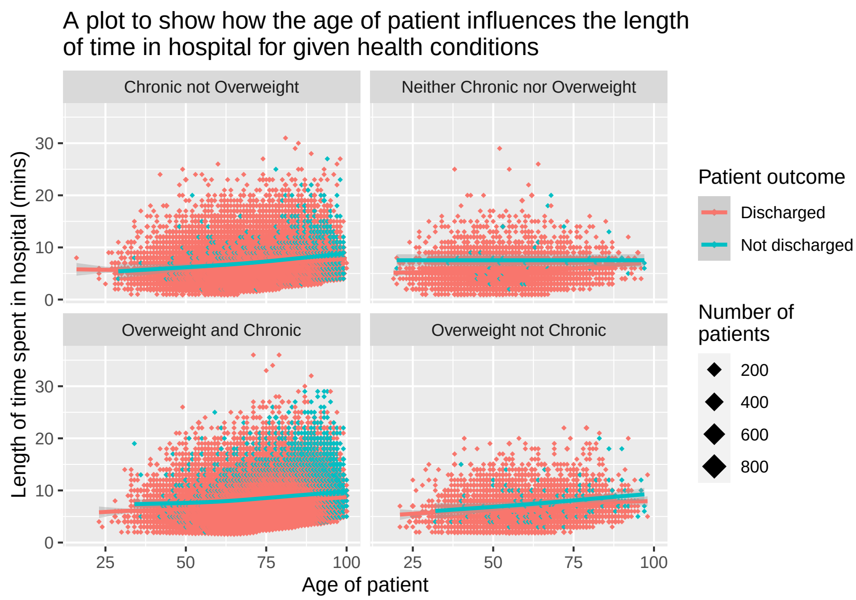
Considering health condition factors, such as age of patient; chronic disease status; and BMI, is there evidence to support that the length of stay in hospital and risk of death could be reduced by better ambulance prioritization?

The question chosen was aimed at further understanding the different health factors of patients, and understanding the how the ambulance service responds to incidences. Chosen to categorize people with a BMI of 25 or above as overweight.

```
Hospital$temp1 = Hospital$Operation
Hospital=Hospital %>% mutate(Operation =
           case_when(BMI >24 & Chronic == 1 ~ 'Overweight and Chronic',
                     BMI >24 & Chronic == 0 ~ 'Overweight not Chronic',
                     Chronic == 1 & BMI <25 ~ 'Chronic not Overweight',
                     Chronic == 0 & BMI <25 ~
                       'Neither Chronic nor Overweight' )) %>%
  rename('Operation' = temp1, 'Health conditions' = Operation )
Hospital %>% group_by(`Health conditions`) %>% summarise('Risk of Death' = mean(Outcome),
    'Mean length of stay' = mean(Length))
```

```
## # A tibble: 4 x 3
##   `Health conditions`            `Risk of Death` `Mean length of stay`
##   <chr>                                    <dbl>                 <dbl>
## 1 Chronic not Overweight                 0.0245                   6.44
## 2 Neither Chronic nor Overweight         0.00916                  5.49
## 3 Overweight and Chronic                 0.0674                   7.58
## 4 Overweight not Chronic                 0.0218                   6.42
```

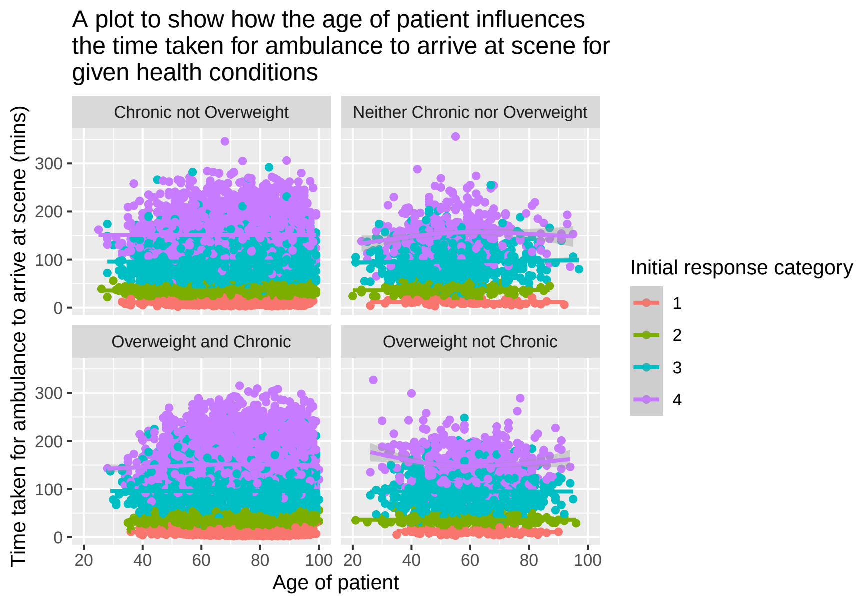
```
Hospital = Hospital %>% mutate(Outcome =
           case_when(Outcome == 1 ~'Not discharged', Outcome == 0 ~ 'Discharged'))
Hospital %>%
  ggplot(aes(y = Length, x = Age, color = Outcome)) + geom_count(shape = 18) +
  facet_wrap(~`Health conditions`) +
  scale_size_continuous(name = 'Number of \npatients') + geom_smooth() +
  labs(x='Age of patient', y='Length of time spent in hospital (mins)',
  title = 'A plot to show how the age of patient influences the length
of time in hospital for given health conditions', color= 'Patient outcome')
```



The above graph and health conditions table show that patient's age, BMI, and chronic disease status effect the length of stay in hospital/health outcome of the patient.

- Patients in `Chronic not Overweight` or in `Overweight and Chronic` tend to spend longer in hospital than patients without these conditions. The length spent in hospital for older people is greater than for younger people with these conditions. This contrasts `Overweight not Chronic` and `Neither Chronic nor Overweight`. As age increases, patients in `Overweight not Chronic` have a small increase in the length of stay in hospital, whereas for patients in `Overweight not Chronic`, the length of stay in hospital says the same.
- There are a significantly higher number of patients who are both overweight and have chronic diseases. These patients also spent the longest time in hospital on average.
- The largest proportion of deaths seen is for patients with `Overweight and Chronic`. These patients have a risk of death 740% higher than patients in `Neither Chronic nor Overweight`. Also, a large number of patients who died in the `Overweight and Chronic` category were elderly; observe that this is similar for the `Chronic not Overweight` category. Whereas the deaths for patients in the `Neither Chronic nor Overweight` category have a fair spread across ages.

```
Utopia$temp1 = Utopia$Operation
Utopia=Utopia %>% mutate(Operation =
           case_when(BMI >24 & Chronic == 1 ~ 'Overweight and Chronic',
                     BMI >24 & Chronic == 0 ~ 'Overweight not Chronic',
                     Chronic == 1 & BMI <25 ~ 'Chronic not Overweight',
                     Chronic == 0 & BMI <25 ~
                       'Neither Chronic nor Overweight' )) %>%
  rename('Operation' = temp1, 'Health conditions' = Operation )
Utopia %>% drop_na(Arrived_Hospital) %>%
  ggplot(aes(y = `Time difference`, x = Age ,color = Initial_Category)) +
  geom_point() + geom_smooth() + facet_wrap(~`Health conditions`) +
  labs(x='Age of patient', y='Time taken for ambulance to arrive at scene (mins)',
  title = 'A plot to show how the age of patient influences
the time taken for ambulance to arrive at scene for
given health conditions', color= 'Initial response category')
```



The second graph shows that Utopia's ambulance services do not consider health factors, specifically, patient age; chronic status; and overweight status, when choosing the initial ambulance response categories 1, 2 and 3 (life threatening (1) to non-urgent (4)).

- The LOESS curves for the initial ambulance response categories 1, 2 and 3 are all parallel, and have a gradient very close to 0. This implies that age does not have an impact on the ambulance arrival time at the scene for these three categories.

- All the LOESS curves for the initial ambulance response categories 1, 2 and 3 have a gradient close to 0, and have the same ambulance arrival time for respective health conditions. This implies that chronic status, and whether the patient is overweight, do not impact the ambulance arrival time at the scene.
- For initial ambulance response category 4, the gradient of the LOESS line is not approximately 0. This implies that age was likely considered when choosing this category. The time it took for the ambulance to arrive at the scene was also not the same for each of the heath conditions. In non-urgent situations, more time is available to ask more questions about the patients medical background than in a life-threatening situation. Therefore, it is likely that these health factors were considered when choosing this initial ambulance response category.

From question 2c, the ambulance arrival time at the scene is faster for patients who have a higher risk of death. From question 2b, patients who have a lower response category have a higher risk of death. Therefore, with aid of the information above (bullet points), the initial ambulance response categories 1, 2 and 3 are chosen without knowledge of patients chronic status, overweight status and age. However, it was shown that the patients chronic status, overweight status and age, effect the length of stay in hospital/health outcome of the patient. This suggests that ambulances could certainly be prioritized more effectively with more information about the patient, specifically, chronic disease status, age, and overweight status. This information could be quickly obtained during the initial call using concise questions. This would almost certainly reduce the risk of death/length of stay in hospital. This is because with a more accurate initial ambulance response category, Utopia's ambulance services will be able to prioritize patients better, hence reducing risk of death/length of stay in hospital for the average patient.

e) *Write a short (two paragraphs) report about the findings of your analysis in parts a-d. The report should be readable for people without data science knowledge. Make it sound interesting and state possible recommendations that may be of interest to Utopia's health department.* **[7 marks]**

During September 2021, at least 67172 calls were made to the health department of Utopia's ambulance services. 80% of these calls lead to the patient arriving at Utopia's hospital. On a typical day, the time of day with the highest frequency of calls made to the ambulance services was 16:00-18:00. The ambulance service was significantly quicker at arriving at the scene to life-threatening incidences, where the risk of death is high, than non-urgent incidences, where the risk of death is low. This suggests that the ambulance service department at Utopia performed well in responding to patients who needed urgent care. Patients who experience life-threatening incidences, tend to spend a longer time in hospital than patients experiencing less urgent incidences. General health factors, precisely, the patient age; chronic disease status; and overweight status, significantly affected the length of stay in hospital/health outcome. Patients who were overweight, and had a chronic disease, had a risk of death 740% higher than patients who were not overweight, and had no chronic diseases.

The ambulance service department does not seem to consider the patient's age; chronic disease status; and overweight status, when sending ambulances to the scene of the incident. The report found evidence to suggest that allowing these health factors to influence the initial ambulance response category, could reduce the length of stay/risk of death. A recommendation to Utopia's ambulance service department would be to ask concise questions, if possible, about the patients chronic status, age, and overweight status, during the initial call. Another recommendation would be to strategically allocate shift patterns, so that more staff are working at times where there are a high frequency of calls. This would allow Utopia's health department to provide a better service to the average patient. These recommendations would certainly be of use to the health department of Utopia; they would definitely increase the performance of Utopia's ambulance services.