

# WEEK 7-8 EXPONENTIAL FAMILY & GLM

Shadan Khan

2024-09-16

###1. Define a string including your name, unit name, task name; and print

```
my_information <- "Name: SHADAN KHAN, Unit: SIT741 , Task: Data Analysis in R in GLM"
print(my_information)
```

```
## [1] "Name: SHADAN KHAN, Unit: SIT741 , Task: Data Analysis in R in GLM"
```

###2. Load the same abalone data set as for the previous pass task.

*#Loading the Abalone Data Set*

```
abalone_data <- read.csv("https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data")
colnames(abalone_data) <- c("Sex", "Length" , "Diameter" , "Height" ,"Whole_Weight", "Shucked_Weight", "Viscera_Weight", "Shell_Weight")
###Converting Rings To Age
abalone_data$Age <- abalone_data$Rings + 1.5
```

###3. Build a linear model using lm() that predicts the age (V9) from two of the numeric variables (V2-V8).

*#Showcasing Linear model predicting Age from Two numerical variables i.e. (Length & Whole\_Weight)*

```
linear_model <- lm(Age ~ Length + Whole_Weight, data = abalone_data)
summary(linear_model)
```

```
##
## Call:
## lm(formula = Age ~ Length + Whole_Weight, data = abalone_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.9976 -1.6747 -0.7428  0.9122 16.3478
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.9313     0.3156   15.63 < 2e-16 ***
## Length       10.5824     0.9071   11.67 < 2e-16 ***
## Whole_Weight   1.1550     0.2221    5.20 2.09e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.67 on 4174 degrees of freedom
## Multiple R-squared:  0.3144, Adjusted R-squared:  0.314
## F-statistic: 956.9 on 2 and 4174 DF,  p-value: < 2.2e-16
```

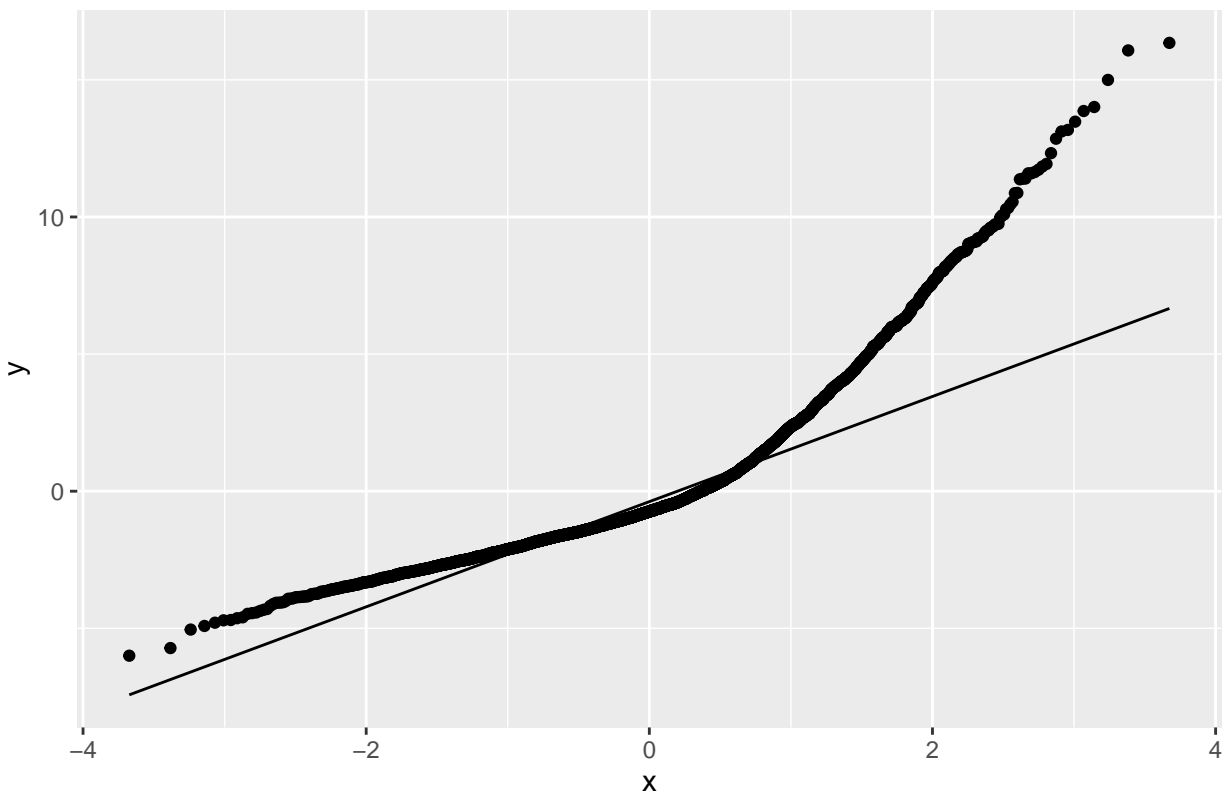
###4. Use the broom and ggplot packages to check the QQ-plot, residuals (for the whole ###model, as well as for each predictor variable), and high leverage points (and whether ###or not any outliers should be excluded from the model).

```
#We load required packages
library(broom)
library(ggplot2)

#We will check residuals & QQ-Plot
model_diagram <- augment(linear_model)

#We will make QQ-Plot
ggplot(model_diagram, aes(sample = .resid)) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("QQ Plot for the Residuals")
```

QQ Plot for the Residuals

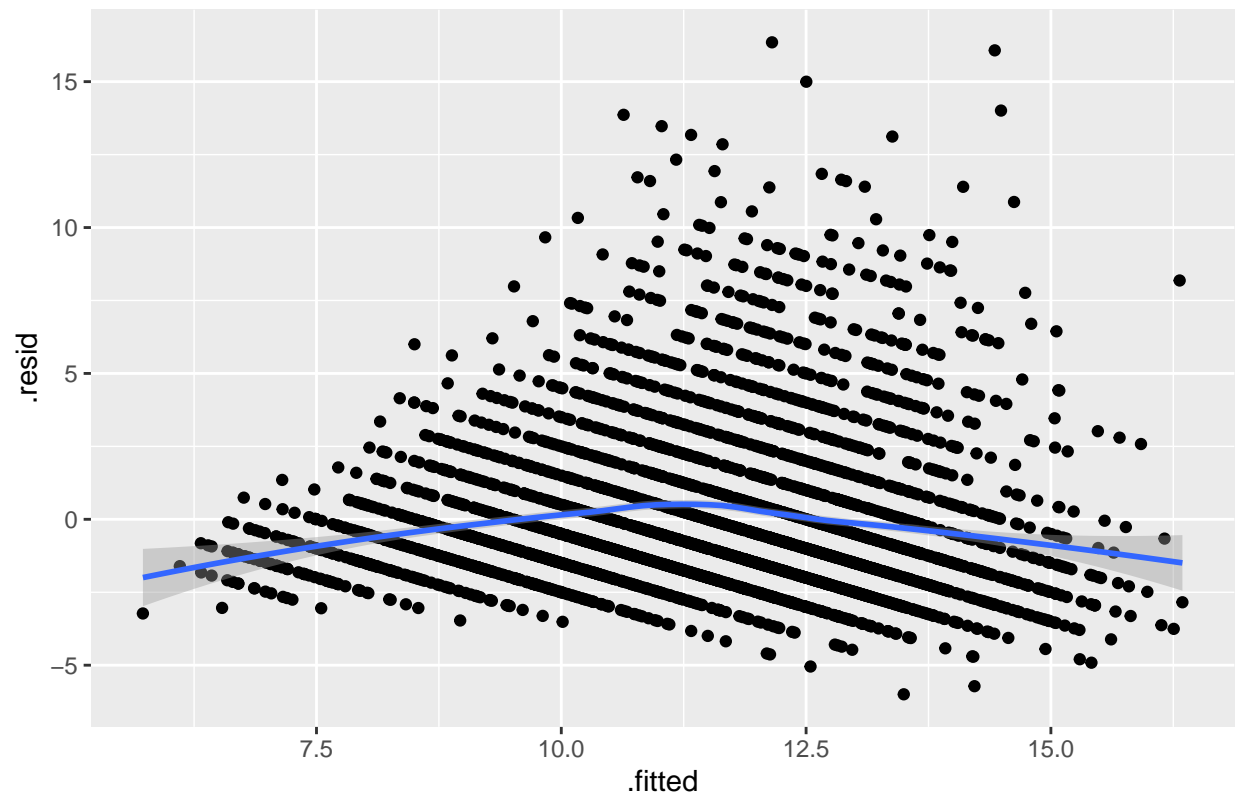


```
#We will showcase residual plot for whole model

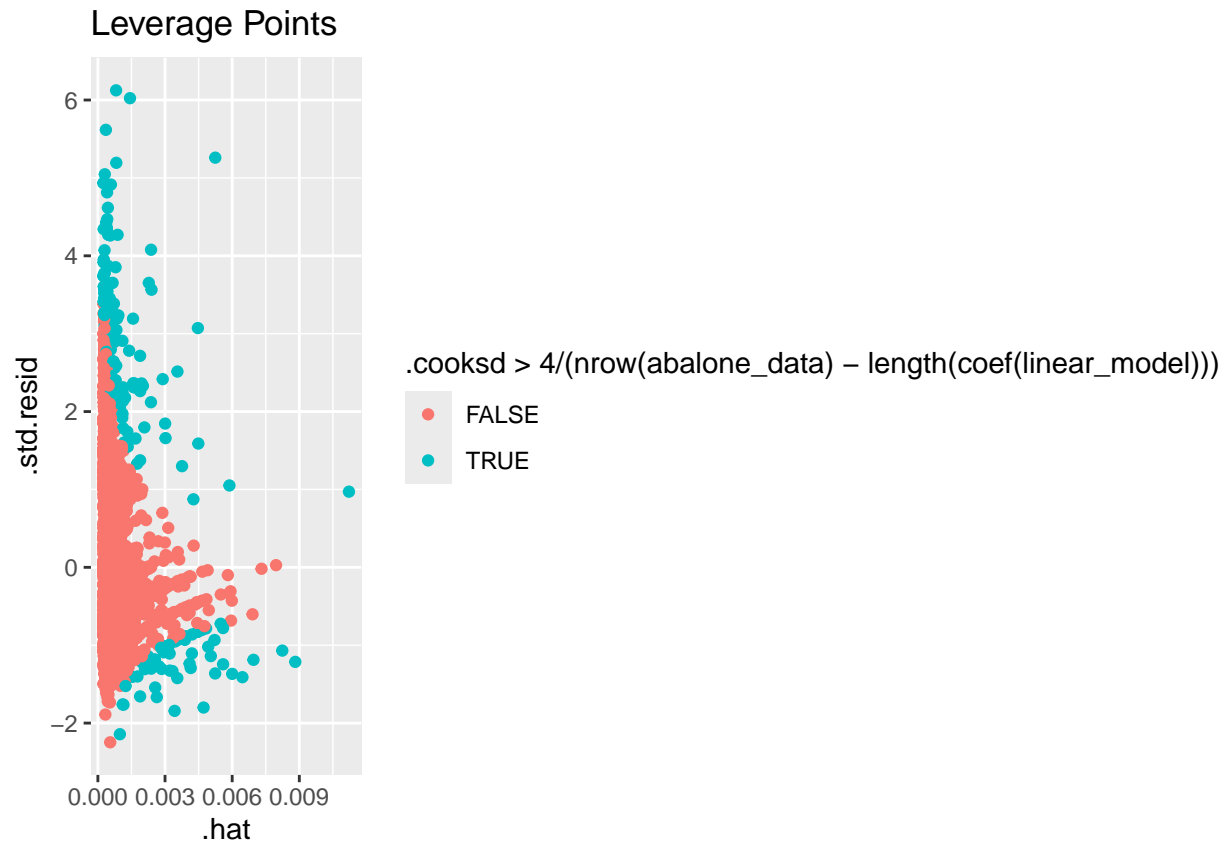
ggplot(model_diagram, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_smooth(method = "loess") +
  ggtitle("Residuals vs Fitted")
```

## 'geom\_smooth()' using formula = 'y ~ x'

Residuals vs Fitted



```
#We will check for high leverage points
ggplot(model_diagram, aes(x= .hat, y=.std.resid)) +
  geom_point(aes(color = .cooksdb > 4/(nrow(abalone_data)- length(coef(linear_model))))) +
  ggtitle("Leverage Points")
```



###5. Fit a generalised linear model using `glm()` with an additional predictor variable ###compared to what you have in step 3. Comment briefly on differences of fit between ###the models.

*##First determining GLM with one additional predictor variable (Diameter)*

```
glm_model <- glm(Age ~ Length + Whole_Weight + Diameter , family = gaussian(), data = abalone_data)
summary(glm_model)
```

```
##
## Call:
## glm(formula = Age ~ Length + Whole_Weight + Diameter, family = gaussian(),
##      data = abalone_data)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.9120     0.3107  15.810 < 2e-16 ***
## Length       -11.8042     2.1341  -5.531 3.37e-08 ***
## Whole_Weight   0.6345     0.2233   2.841 0.00451 **
## Diameter      29.8645     2.5857  11.550 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 6.911407)
##
##      Null deviance: 43411  on 4176  degrees of freedom
## Residual deviance: 28841  on 4173  degrees of freedom
```

```
## AIC: 19935
##
## Number of Fisher Scoring iterations: 2
```

*#Note: Here we can compare the summary outputs of 'lm\_model' and 'glm\_model' to discuss the fit , resid*

*#Now comparing AIC of both models*  
AIC (linear\_model , glm\_model)

```
##           df      AIC
## linear_model  4 20064.12
## glm_model     5 19934.68
```

##6. Download the hospital emergency department admissions dataset.

```
heda <- read.csv("https://data.gov.au/data/dataset/6bfec5ea-207e-4d67-8965-c7e72290844b/resource/33d849")
summary(heda)
```

```
##           X           Royal.Perth.Hospital      X.1           X.2
## Length:366      Length:366      Length:366      Length:366
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##           X.3           X.4           X.5           X.6
## Length:366      Length:366      Length:366      Length:366
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
## Fremantle.Hospital X.7           X.8           X.9
## Length:366      Length:366      Length:366      Length:366
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##           X.10          X.11          X.12
## Length:366      Length:366      Length:366
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
## Princess.Margaret.Hospital.For.Children X.13          X.14
## Length:366      Length:366      Length:366
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##           X.15          X.16          X.17          X.18
## Length:366      Length:366      Length:366      Length:366
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
## King.Edward.Memorial.Hospital.For.Women X.19          X.20
## Length:366      Length:366      Length:366
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##           X.21          X.22          X.23          X.24
## Length:366      Length:366      Length:366      Length:366
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
## Sir.Charles.Gairdner.Hospital X.25          X.26
```

```

## Length:366          Length:366          Length:366
## Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character
##      X.27            X.28            X.29            X.30
## Length:366          Length:366          Length:366          Length:366
## Class :character    Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character     Mode :character
## Armadale.Kelmscott.District.Memorial.Hospital      X.31
## Length:366          Length:366
## Class :character    Class :character
## Mode :character     Mode :character
##      X.32            X.33            X.34            X.35
## Length:366          Length:366          Length:366          Length:366
## Class :character    Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character     Mode :character
##      X.36            Swan.District.Hospital      X.37
## Length:366          Length:366          Length:366
## Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character
##      X.38            X.39            X.40            X.41
## Length:366          Length:366          Length:366          Length:366
## Class :character    Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character     Mode :character
##      X.42            Rockingham.General.Hospital      X.43
## Length:366          Length:366          Length:366
## Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character
##      X.44            X.45            X.46            X.47
## Length:366          Length:366          Length:366          Length:366
## Class :character    Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character     Mode :character
##      X.48            Joondalup.Health.Campus      X.49
## Length:366          Length:366          Length:366
## Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character
##      X.50            X.51            X.52            X.53
## Length:366          Length:366          Length:366          Length:366
## Class :character    Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character     Mode :character
##      X.54
## Length:366
## Class :character
## Mode :character

```

```
head(heda)
```

```

##      X Royal.Perth.Hospital      X.1  X.2  X.3  X.4  X.5  X.6
## 1      Date      Attendance Admissions Tri_1 Tri_2 Tri_3 Tri_4 Tri_5
## 2 01-JUL-2013      235      99      8      33      89      85      20
## 3 02-JUL-2013      209      97      N/A     41      73      80      14
## 4 03-JUL-2013      204      84      7      40      72      79      6
## 5 04-JUL-2013      199     106      3      37      73      70      15
## 6 05-JUL-2013      193      96      4      40      76      62      11
## Fremantle.Hospital      X.7  X.8  X.9  X.10 X.11 X.12

```

##	1	Attendance		Admissions	Tri_1	Tri_2	Tri_3	Tri_4	Tri_5					
##	2	155		70	N/A	25	67	54	9					
##	3	145		56	N/A	22	51	47	24					
##	4	118		60	N/A	24	37	43	12					
##	5	125		61	N/A	21	47	42	14					
##	6	136		58	4	23	51	45	13					
##	Princess.Margaret.Hospital.For.Children						X.13	X.14	X.15	X.16	X.17			
##	1	Attendance		Admissions	Tri_1	Tri_2	Tri_3	Tri_4						
##	2	252			59	N/A	13	75	159					
##	3	219			47	N/A	14	61	139					
##	4	186			31	N/A	11	53	120					
##	5	192			41	N/A	8	57	126					
##	6	201			42	N/A	14	53	130					
##	X.18 King.Edward.Memorial.Hospital.For.Women						X.19	X.20	X.21	X.22				
##	1	Tri_5		Attendance		Admissions	Tri_1	Tri_2	Tri_3					
##	2	4		52			6	N/A	N/A	7				
##	3	4		43			7	N/A	N/A	N/A				
##	4	N/A		55			5	N/A	N/A	4				
##	5	N/A		42			6	N/A	N/A	7				
##	6	N/A		64			3	N/A	N/A	5				
##	X.23		X.24	Sir.Charles.Gairdner.Hospital		X.25		X.26	X.27	X.28	X.29			
##	1	Tri_4		Tri_5		Attendance		Admissions	Tri_1	Tri_2	Tri_3	Tri_4		
##	2	20		25		209		109	N/A	42	108	47		
##	3	25		17		184		112	N/A	43	73	61		
##	4	23		28		171		102	N/A	49	69	48		
##	5	21		14		153		89	6	35	57	46		
##	6	36		22		167		117	7	39	70	46		
##	X.30 Armadale.Kelmscott.District.Memorial.Hospital						X.31		X.32	X.33				
##	1	Tri_5		Attendance		Admissions	Tri_1	Tri_2						
##	2	11		166			19	N/A	16					
##	3	5		175			25	N/A	26					
##	4	3		145			11	N/A	19					
##	5	9		144			15	N/A	15					
##	6	5		147			18	N/A	17					
##	X.34		X.35	X.36 Swan.District.Hospital		X.37		X.38	X.39	X.40	X.41			
##	1	Tri_3		Tri_4		Tri_5		Attendance		Admissions	Tri_1	Tri_2	Tri_3	Tri_4
##	2	62		79		9		133		17	N/A	29	59	42
##	3	73		55		19		129		25	N/A	25	57	43
##	4	62		60		4		126		7	3	24	42	55
##	5	48		69		12		108		19	N/A	14	45	43
##	6	59		63		7		122		24	3	15	61	37
##	X.42 Rockingham.General.Hospital						X.43		X.44	X.45	X.46	X.47	X.48	
##	1	Tri_5		Attendance		Admissions	Tri_1	Tri_2	Tri_3	Tri_4	Tri_5			
##	2	N/A		155		10	N/A	12	51	81	11			
##	3	4		145		24	N/A	26	45	65	8			
##	4	N/A		151		14	N/A	10	50	87	3			
##	5	6		129		15	N/A	24	45	55	4			
##	6	6		122		12	N/A	19	46	53	4			
##	Joondalup.Health.Campus						X.49		X.50	X.51	X.52	X.53		
##	1	Attendance		Admissions	Tri_1	Tri_2	Tri_3	Tri_4						
##	2	267		73	N/A	27	75	151						
##	3	241		81	N/A	23	78	133						
##	4	213		67	N/A	29	66	99						
##	5	227		72	5	26	68	117						

```
## 6                229          71   N/A    37    80   106
##
## 1
## 2              12
## 3              7
## 4             17
## 5             11
## 6 4
```

###7. Plot the attendance and admissions data over time (as a single graph) for one of the ###hospitals using ggplot. You will need to make some adjustments to the dataset in ###order to do this. You can use, e.g., `dmy("10-JUL-2024")` in the `lubridate` package to ###convert it to a 'date' datatype.

```
#loading lubridate
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

```
#Converting data column to data type
heda$Date <- dmy(heda$X)
```

```
## Warning: 1 failed to parse.
```

```
#Now we will filter data for specific hospital
colnames(heda)
```

```
## [1] "X"
## [2] "Royal.Perth.Hospital"
## [3] "X.1"
## [4] "X.2"
## [5] "X.3"
## [6] "X.4"
## [7] "X.5"
## [8] "X.6"
## [9] "Fremantle.Hospital"
## [10] "X.7"
## [11] "X.8"
## [12] "X.9"
## [13] "X.10"
## [14] "X.11"
## [15] "X.12"
## [16] "Princess.Margaret.Hospital.For.Children"
## [17] "X.13"
## [18] "X.14"
## [19] "X.15"
## [20] "X.16"
## [21] "X.17"
```



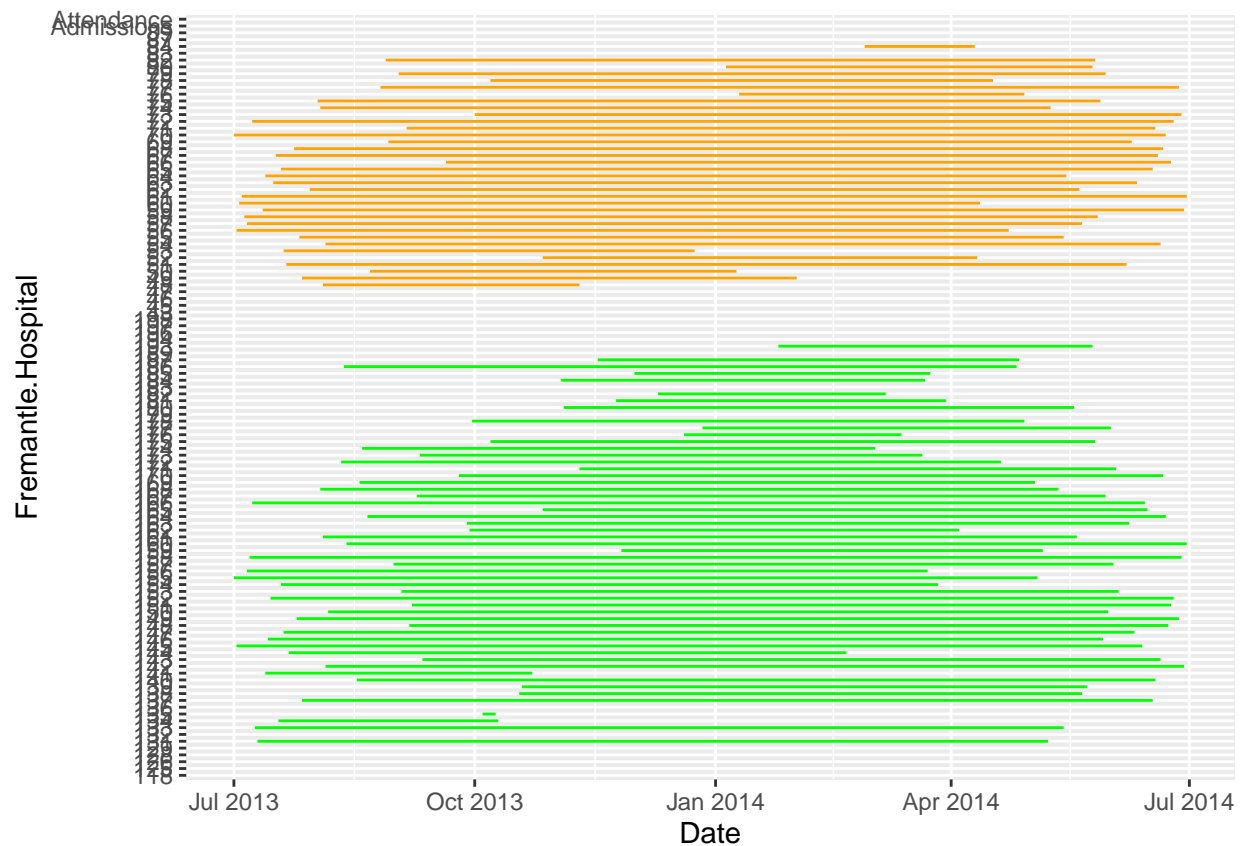
```
## [22] "X.18"
## [23] "King.Edward.Memorial.Hospital.For.Women"
## [24] "X.19"
## [25] "X.20"
## [26] "X.21"
## [27] "X.22"
## [28] "X.23"
## [29] "X.24"
## [30] "Sir.Charles.Gairdner.Hospital"
## [31] "X.25"
## [32] "X.26"
## [33] "X.27"
## [34] "X.28"
## [35] "X.29"
## [36] "X.30"
## [37] "Armadale.Kelmscott.District.Memorial.Hospital"
## [38] "X.31"
## [39] "X.32"
## [40] "X.33"
## [41] "X.34"
## [42] "X.35"
## [43] "X.36"
## [44] "Swan.District.Hospital"
## [45] "X.37"
## [46] "X.38"
## [47] "X.39"
## [48] "X.40"
## [49] "X.41"
## [50] "X.42"
## [51] "Rockingham.General.Hospital"
## [52] "X.43"
## [53] "X.44"
## [54] "X.45"
## [55] "X.46"
## [56] "X.47"
## [57] "X.48"
## [58] "Joondalup.Health.Campus"
## [59] "X.49"
## [60] "X.50"
## [61] "X.51"
## [62] "X.52"
## [63] "X.53"
## [64] "X.54"
## [65] "Date"
```

*#now by plotting admissions and attendance over time*

```
ggplot(heda, aes(x = Date)) +
  geom_line(aes(y= Fremantle.Hospital),color = "green") +
  geom_line(aes(y= X.7), color= "orange")
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_line()').
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_line()').
```



### 8. Use the `wkday` functions to show the data for a single day of the week and plot the data over time, then compare this to the results for another day.

```
#First we Filter data for the single day of the week (e.g Tuesday)
```

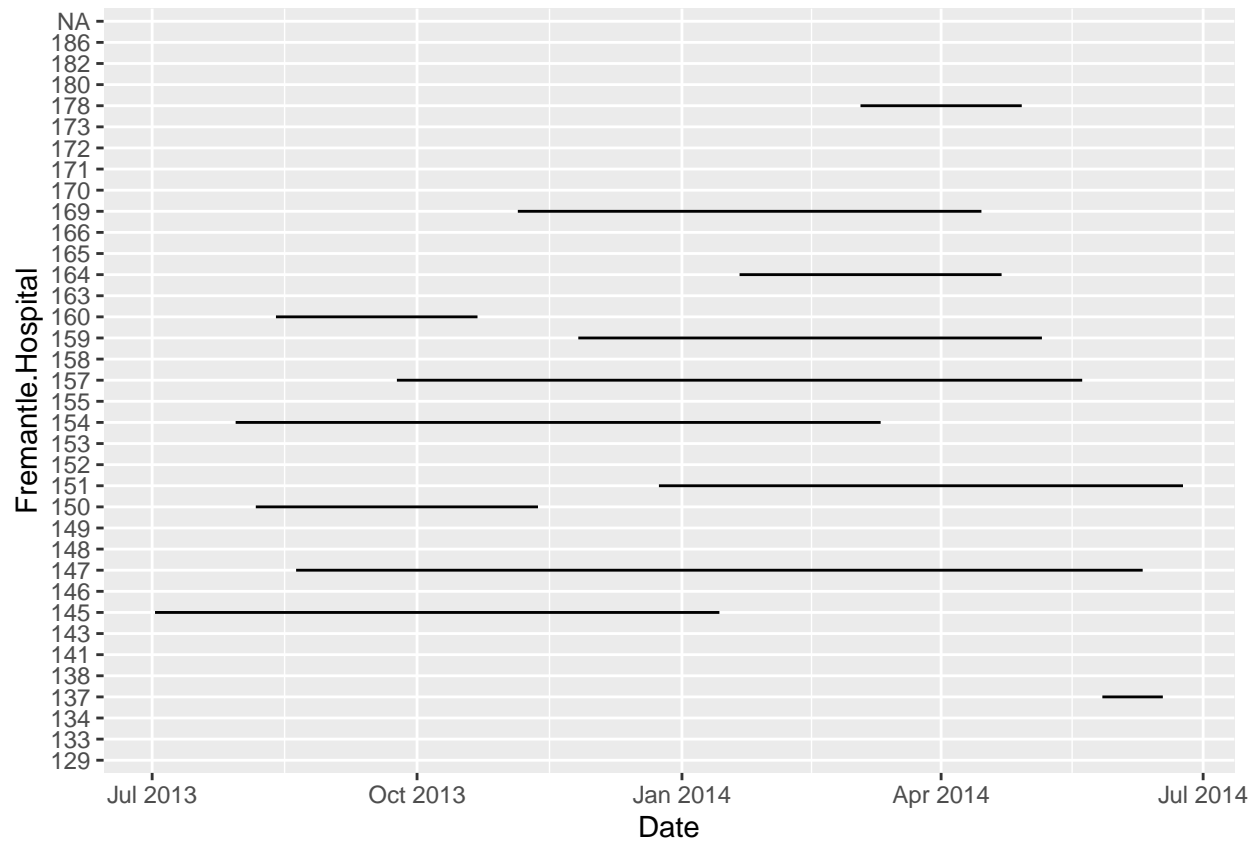
```
library(lubridate)
```

```
tuesday_data <- heda[weekdays(heda$Date) == "Tuesday" , ]
```

```
#Now Plotting Tuesday data over time
```

```
ggplot(tuesday_data, aes(x= Date , y = Fremantle.Hospital)) + geom_line()
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_line()').
```

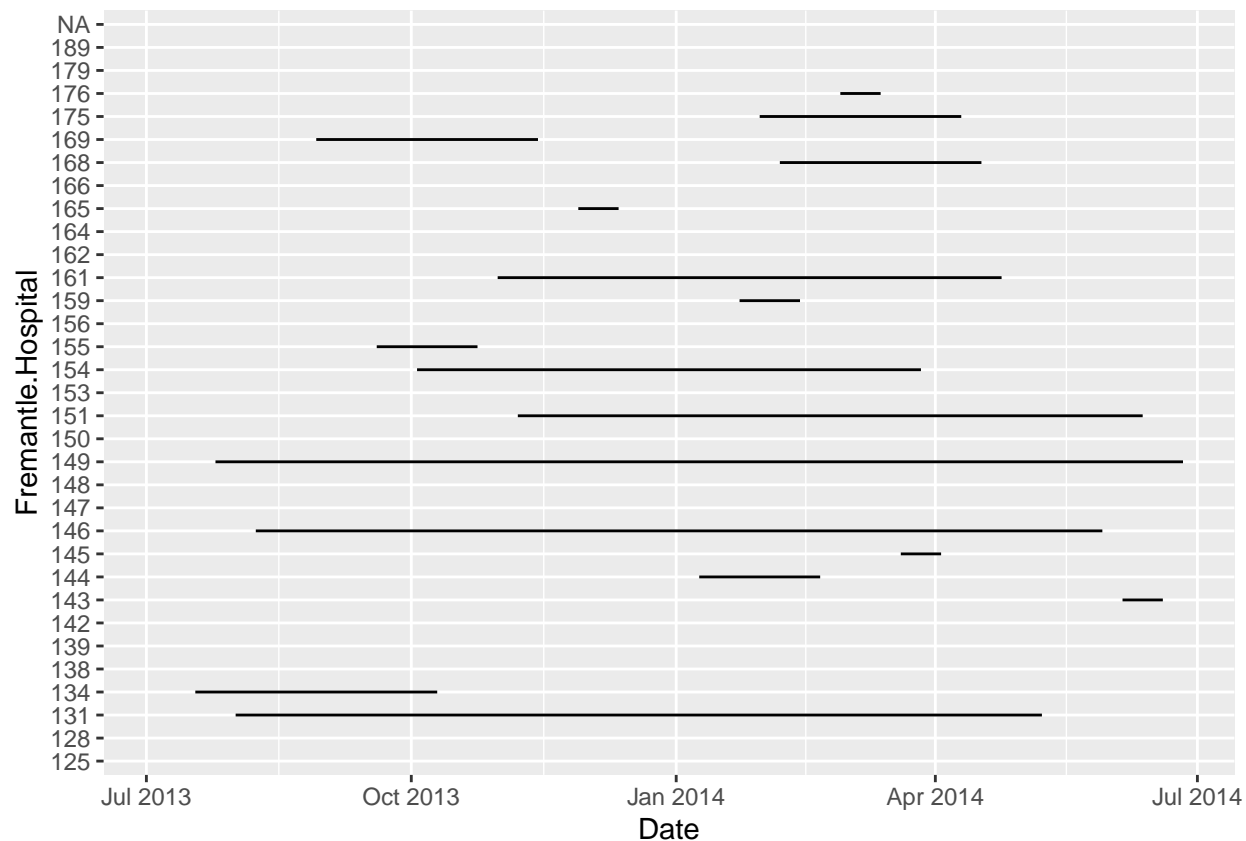


*#Filtering and plotting data for another day*

```
thursday_data <- heda[weekdays(heda$Date) == "Thursday" , ]
```

```
ggplot(thursday_data , aes(x= Date , y = Fremantle.Hospital)) + geom_line()
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_line()').
```



## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

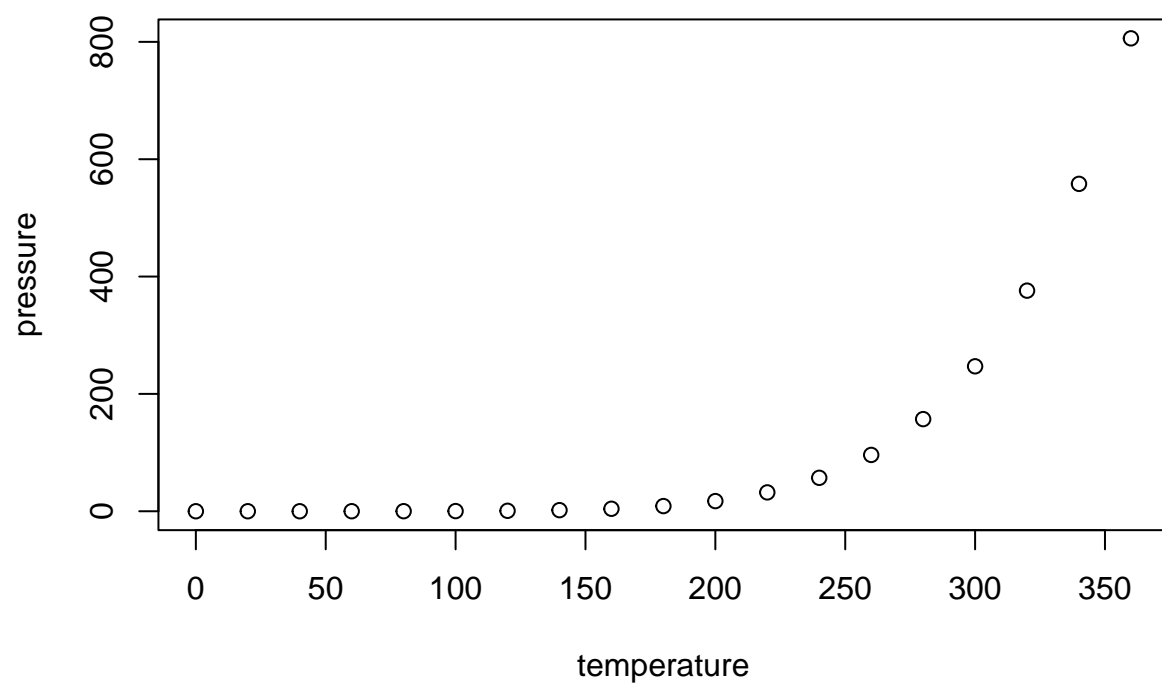
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean    : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.    :120.00
```

## Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.