A1 Individual Report

Data, Insights and Decisions (University of New South Wales)

Executive Summary

1. Relationship between Job Roles and Employee Attrition

A driving factor which may be a cause of negative employee sentiment within Globex's workforce could be due the structure and job roles within the company. Figure 1 showcases the percentage make-up of Globex's staff. These proportions can be analysed with respect to Figure 2 which provides a five-metric summary of the monthly incomes associated with each different position.

Figure 1: Pie Chart of the Job Role Composition at Globex Pharma

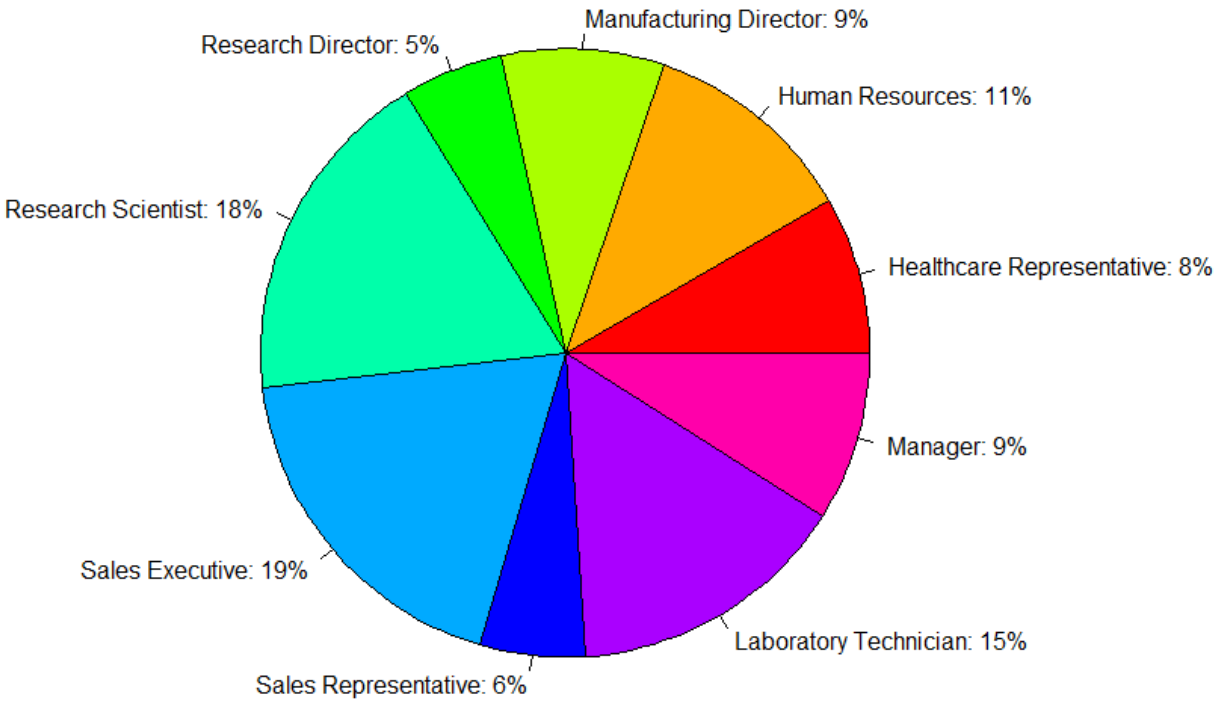


Figure 2: Boxplot - Monthly Income based on Job Role

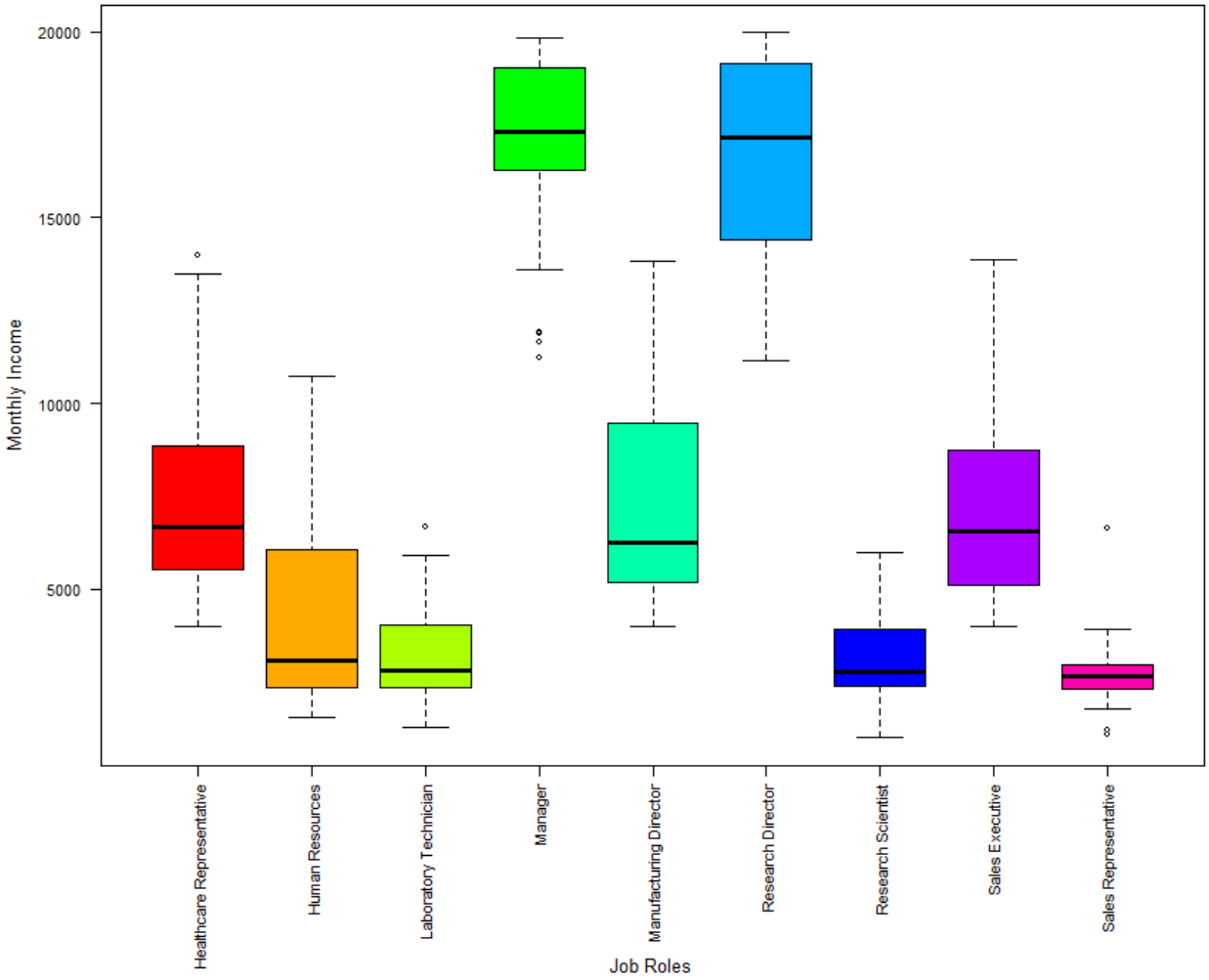


Figure 2 shows that managers and research directors have the highest median monthly incomes whereas employees in human resources, laboratory technicians, research scientists and sales representatives receive the lowest wages.

Using Figure 1 and 2, a factor which may be identified is the income inequality faced by different groups in Globex's workforce. For example, even though there is an almost proportional number of research scientists and sales executives within the company (18% and 19% respectively), there is a difference of $3795.50 between the median incomes of both groups. Moreover, it can be visualised that the interquartile range (IQR) of incomes for sales executives is larger and varies to a greater extent than the incomes for research

scientists and laboratory technicians. This indicates that a majority of employees in technical roles have lower scope in accessing higher wage outcomes than employees in fundamentally non-technical roles such as sales executives and managers. This may contribute to increasing dissatisfaction within the work that employees in these roles do.

Figure 3: Boxplot - Salary Increases based on Job Role

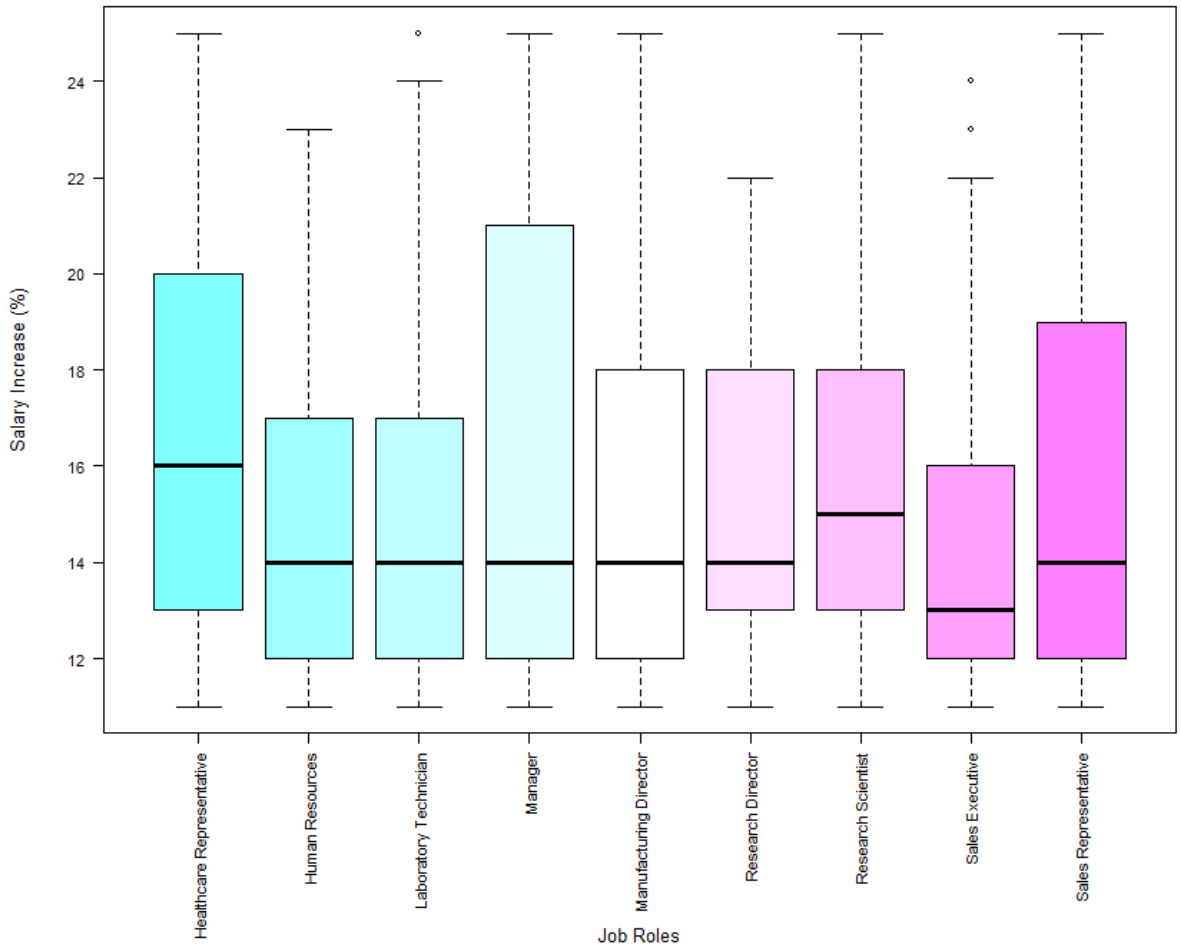


Figure 3 summarises the salary increases experienced by employees in each role, allowing us to examine inconsistencies in Globex's pay structure. Although each job role excluding sales executives last experienced a median salary increase ranging between 14-16%, some roles have a much greater IQR than others (i.e. Managers, Sales Representatives), indicating greater variability about the median. The lack of consistency in pay rises may cause negative sentiment among employee groups with lower variability, potentially leading to higher rates of attrition.

2. Factors affecting Income and relationship with Attrition

Factors deriving from income may be a key driver of increased employee frustration and dissatisfaction, leading to greater worker attrition. Figure 4 includes a scatter plot and line of best fit which create a visual relationship between an employee's total working years and their income.

Figure 4: Scatterplot - Total Working Years vs Monthly Income

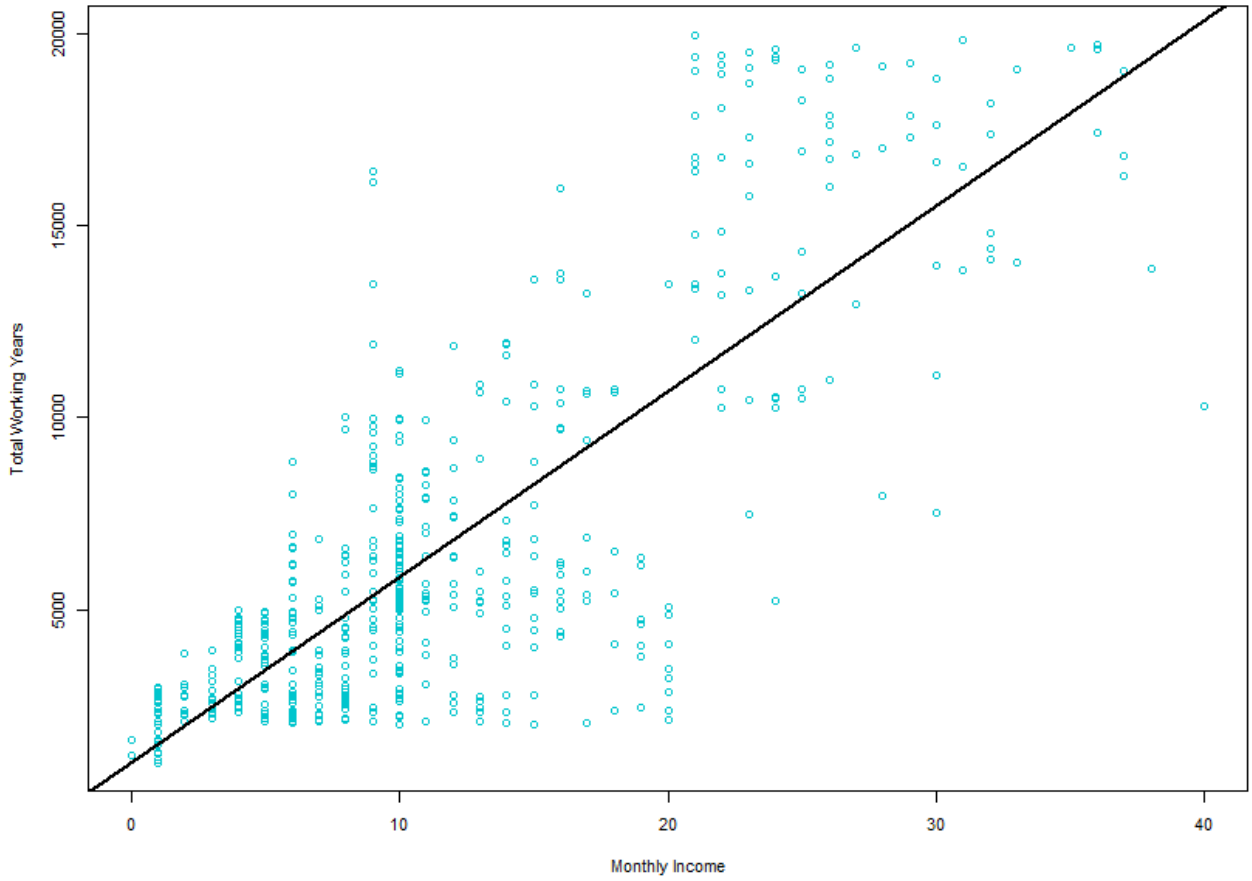
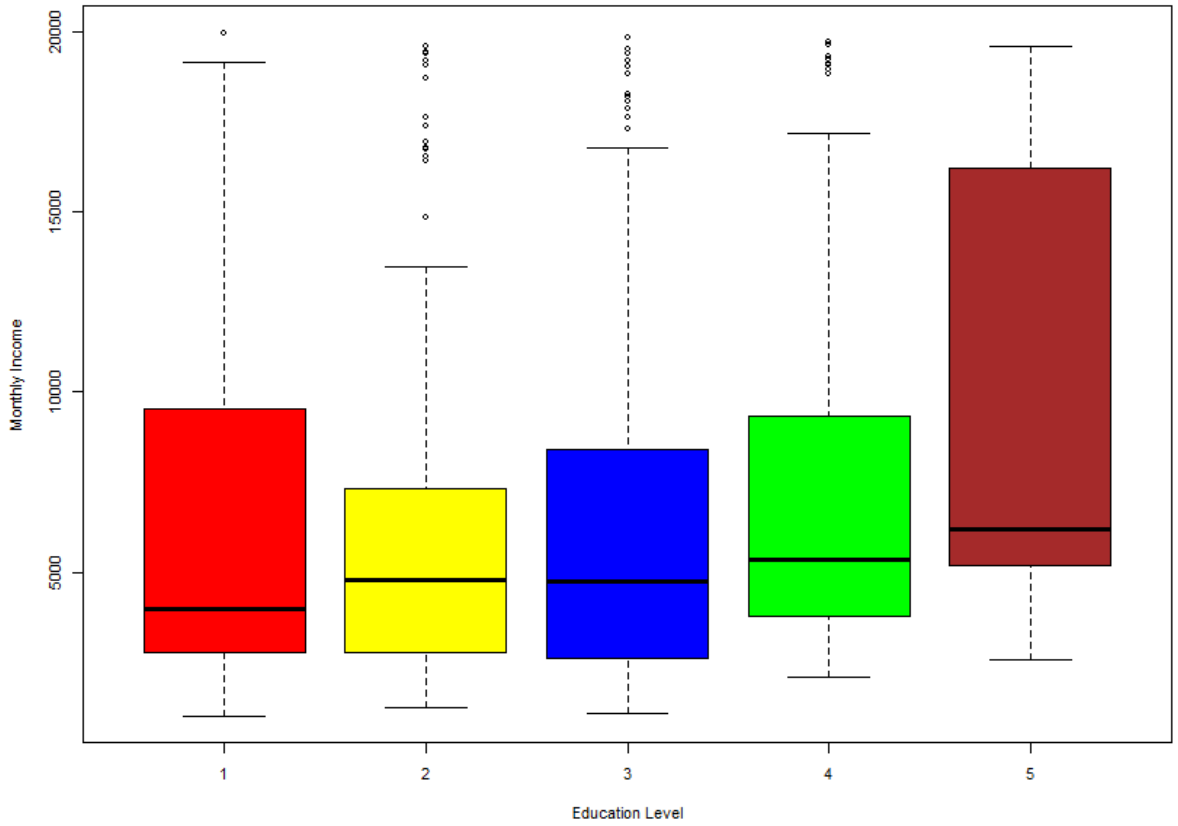


Figure 4 shows a strong positive correlation between the two variables **(R = 0.789),** indicating that workers with greater work experience are more likely to have greater wage outcomes. This may however create dissatisfaction and negative sentiment among employees with a lower number of working years as they may feel restricted in their ability to access higher wages at Globex, leading them to search for different jobs with higher potential incomes while requiring fewer years of work experience.

Figure 5: Boxplot - Monthly Income based on Level of Education



.

Moreover, Figure 5 indicates that employees with higher levels of education have a higher median income level in comparison to workers who only completed high school or a diploma. This can act as a factor causing frustration and higher levels of turnover as employees who may be less educated yet are able to perform tasks equally well may feel less valued by the company and seek other job opportunities.

Figure 6: Boxplot - Job Satisfaction based on Monthly Income

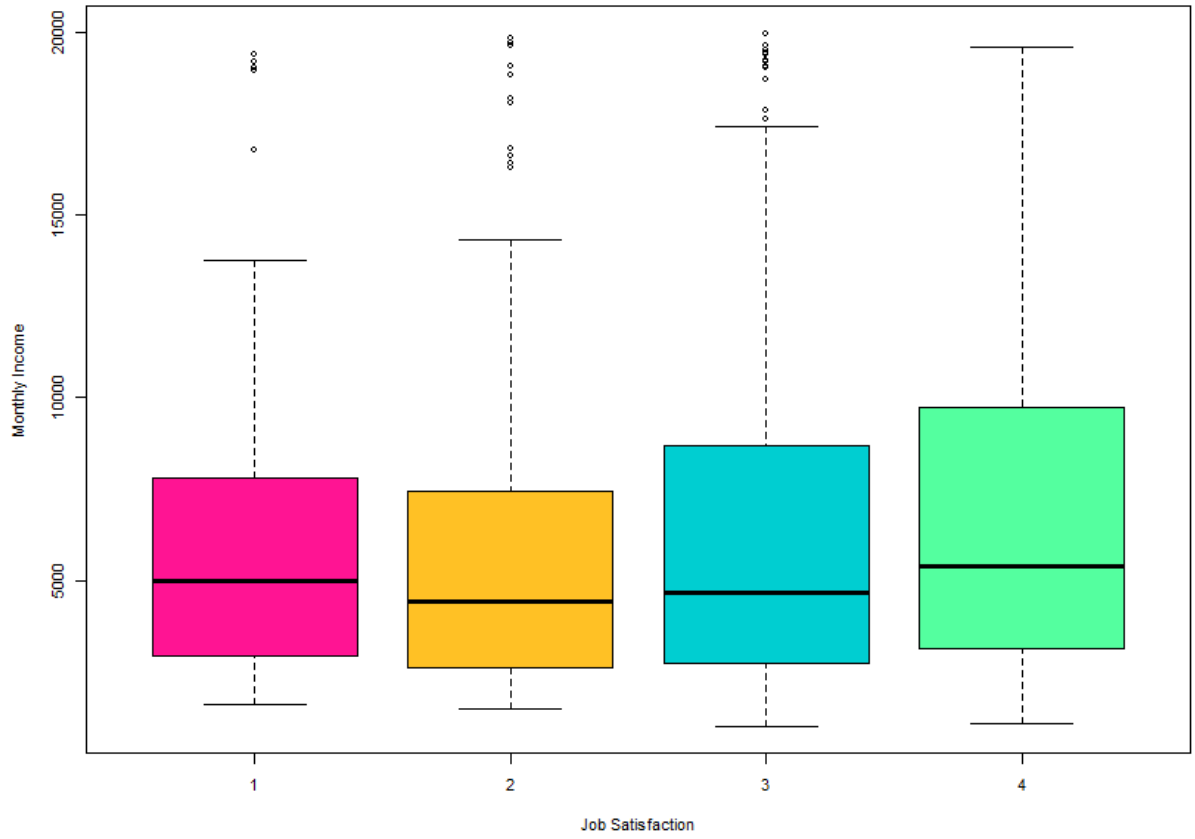


Figure 6 compares income with job satisfaction which should be considered a key variable impacting an individual sentiment. The graph shows that employees who are in Group 4 of Job Satisfaction have a higher median income than employees in Group 2 and 3. However it should also be noted that employees in Group 1-3 have a higher number of top outliers which indicates that higher incomes do not necessarily create greater job satisfaction and other variables including past performance ratings and time since last promotion should be accounted for in determining factors leading to worker dissatisfaction and turnover.

3. Relationship between Employee Age

Finally, another driving factor of employee dissatisfaction and potential cause of higher turnover at Globex Pharma could be due to an imbalance in the demographics of the

company's workforce. These imbalances can be seen in Figure 7 and 8 and the table of descriptive statistics.

Descriptive Statistics	
Mean	37.439
Median	36
Skewness	0.426
Kurtosis	2.474

Figure 7: Employee Age - Histogram vs. Normal PDF

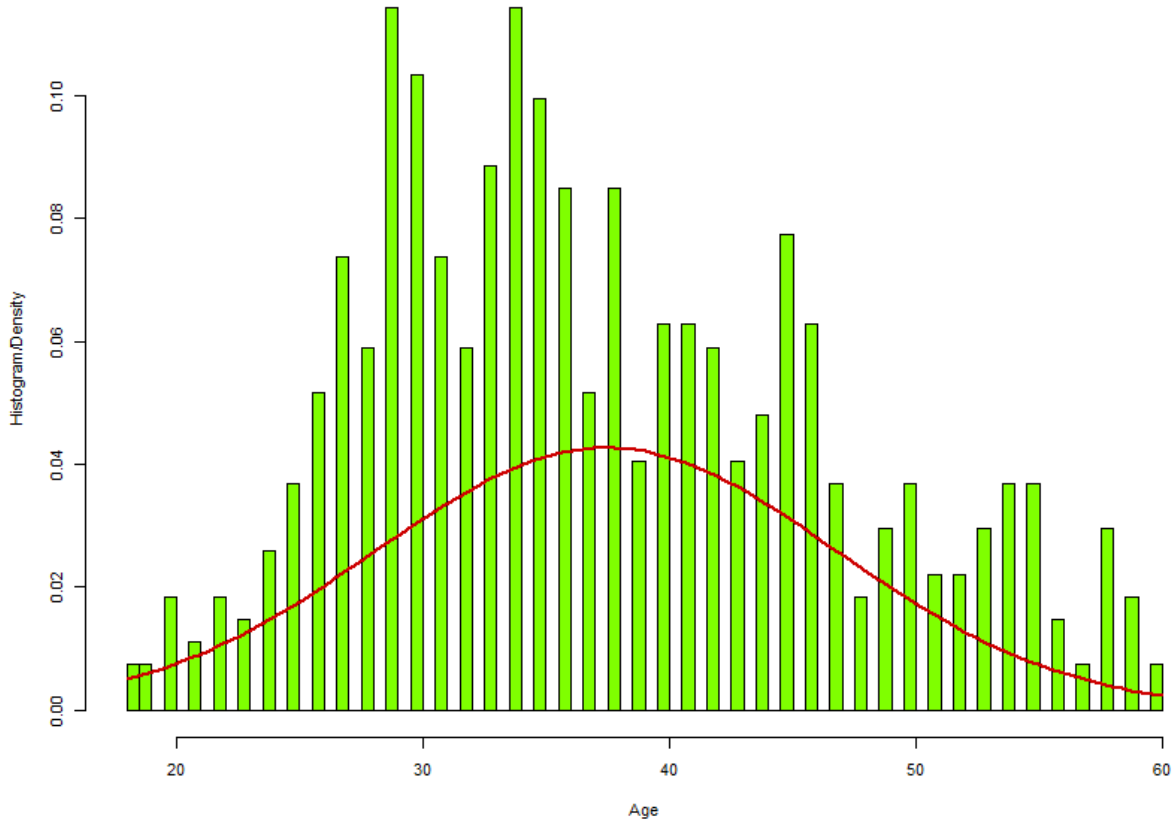


Figure 8: Q-Q Plot - Employee Age

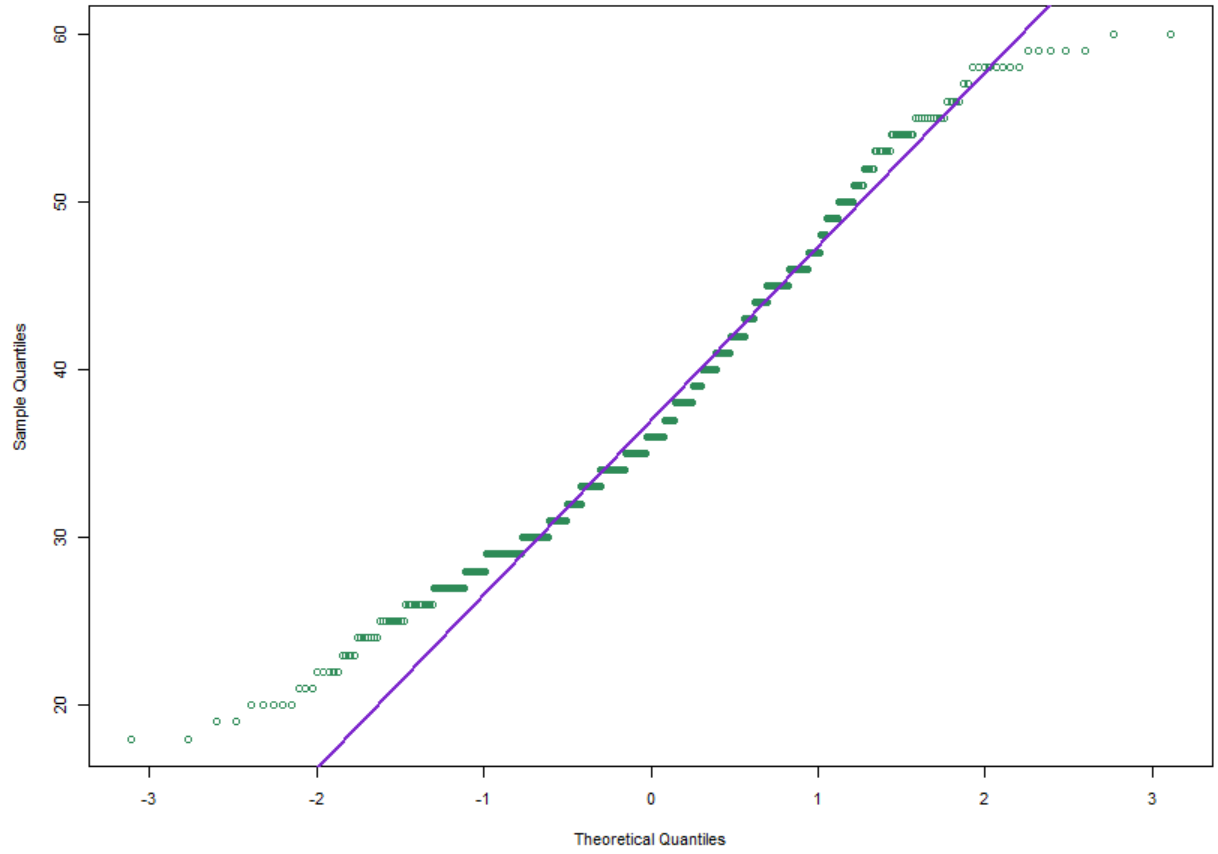


Figure 7 highlights that the ages of employees at Globex are positively skewed in comparison to the histogram, indicating that younger staff members make up a greater proportion of the workforce. This can also be extracted from the Q-Q Plot in Figure 8 in which the values above the Q-Q Line indicate a positive skew. Moreover, the distribution for age is platykurtic since the Kurtosis < 3 and there are values on the left side of the Q-Q Plot above the line of normal distribution. This indicates a very low probability of employees being aged higher or lower than the corresponding sample quintiles.

These findings support Age as a driving factor towards negative sentiment as a greater proportion of younger staff members may cause staff members who are 37+ years old to feel less valued within the company structure. This may further create a stigma of decreased job security due to a potential trend of a younger workforce within Globex,

causing older employees to seek other jobs.

Figure 9: Scatterplot - Monthly Income based on Age

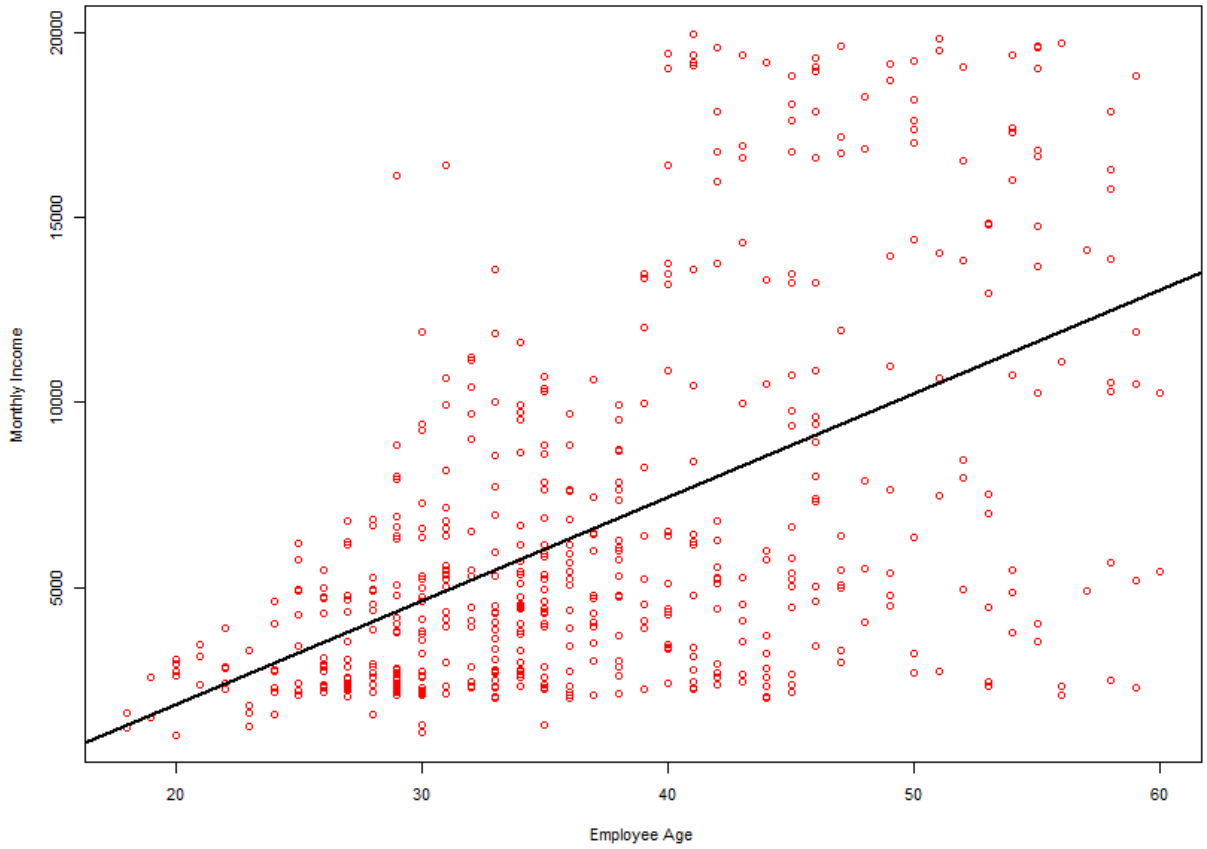


Figure 9 also shows a medium-to-strong linear positive correlation between Age and Monthly Income (**R = 0.528**). This links to our previously discussed relationship between total working years and monthly income as the sentiment that only older members of Globex's staff usually qualify for higher pay. This may drive employee frustration and increased uncertainty in Globex's value of its staff.

Appendix

```r
library(moments)

#Variable 'UData' stands for the usable data extracted from the original 'A1_Dataset.csv'
rows <- c(1:542)
UData <- A1_Dataset[rows,]

#Create variables
JobSatisfaction <- UData[,11]
MonthlyIncome <- UData[,13]
Education <- UData[, 5]
Age <- UData[,1]
TotalWorkingYears <- UData[,19]
SalaryIncrease <- UData[,17]
JobRole <- UData[,10]
WorkLifeBalance <- UData[,21]
DistanceFromHome <- UData[,4]

#FIGURE 1: PIE CHART FOR JOB ROLES
#Step 1 - Create bins for each different type of job role
sum(UData$JobRole == "Healthcare Representative")
sum(UData$JobRole == "Human Resources")
sum(UData$JobRole == "Manufacturing Director")
sum(UData$JobRole == "Research Director")
sum(UData$JobRole == "Research Scientist")
sum(UData$JobRole == "Sales Executive")
sum(UData$JobRole == "Sales Representative")
sum(UData$JobRole == "Laboratory Technician")
sum(UData$JobRole == "Manager")

#Step 2 - Create the Pie Chart
slices <- c(45, 62,47,29,98,101,30,81,49)
JobRoleLabels <- c("Healthcare Representative","Human Resources","Manufacturing Director","Research Director",
                   "Research Scientist","Sales Executive","Sales Representative","Laboratory Technician","Manager")
percentage <- round(slices/sum(slices) * 100)
cpercentage <- paste(JobRoleLabels, ": ", percentage, "%", sep = "")
pie(slices, labels = cpercentage, col = rainbow(length(cpercentage)))
title(main = "Figure 1: Pie Chart of the Job Role Composition at Globex Pharma", cex.main = 1.2)

#Find median incomes of Sales Executives vs Research Scientists
medRS <- median(MonthlyIncome[UData$JobRole == "Research Scientist"]); avgRS
medSE <- median(MonthlyIncome[UData$JobRole == "Sales Executive"]); avgSE
paste("Difference:", medSE - medRS)
```

```r

#FIGURE 2: BOXPLOT - MONTHLY INCOME VS JOB ROLES
par(mar = c(15, 7, 7, 7), cex = 0.7)
boxplot(MonthlyIncome~JobRole, col= rainbow(length(JobRoleLabels)),
        xlab="", ylab="",
        main = "Figure 2: Boxplot - Monthly Income based on Job Role",
        cex.main = 1.2,las =2,data = UData)
title(ylab="Monthly Income",line=4, cex.lab=1.2)
title(xlab = "Job Roles", line=10, cex.lab=1.2)

#FIGURE 3: BOXPLOT - SALARY INCREASE VS JOB ROLES
par(mar = c(16, 8, 8, 8), cex = 0.7)
boxplot(SalaryIncrease~JobRole, col= cm.colors(length(JobRoleLabels)),
        xlab="", ylab="",
        main = "Figure 3: Boxplot - Salary Increases based on Job Role",
        cex.main = 1.2,las =2,data = UData)
title(ylab="Salary Increase (%)",line=4, cex.lab=1.2)
title(xlab = "Job Roles", line=10, cex.lab=1.2)

#FIGURE 4: SCATTERPLOT - TOTAL WORKING YEARS Vs MONTHLY INCOME
plot(TotalWorkingYears, MonthlyIncome,xlab = "Monthly Income", ylab = "Total Working Years",
     main = "Figure 4: Scatterplot - Total Working Years vs Monthly Income", cex.main = 1.2, col = 'turquoise3')
abline(lm(UData$MonthlyIncome~UData$TotalWorkingYears), lwd = 2.5)
cor(TotalWorkingYears, MonthlyIncome)

#FIGURE 5: BOXPLOT - MONTHLY INCOME V EDUCATION
boxplot(MonthlyIncome~Education, col= c("red","yellow","blue", "green", "brown"),
        xlab="Education Level", ylab="Monthly Income",
        main = "Figure 5: Boxplot - Monthly Income based on Level of Education",
        cex.main = 1.2,data = UData)

#FIGURE 6: BOXPLOT - MONTHLY EDUCATION V JOB SATISFACTION
boxplot(MonthlyIncome~JobSatisfaction, col= c("deeppink","goldenrod1","darkturquoise", "seagreen1"),
        xlab="Job Satisfaction", ylab="Monthly Income",
        main = "Figure 6: Boxplot - Job Satisfaction based on Monthly Income",
        cex.main = 1.2,data = UData)
```

```
79
80  #FIGURE 7: HISTOGRAM + NORMAL PDF - EMPLOYEE AGE
81  #Step 1: Find the mean and standard deviation of variable 'Age' as required to create the Normal Distribution.
82  mu <- mean(Age)
83  sigma <- sd(Age)
84
85  #Step 2: Create the Histogram and PDF
86  hist <- hist(Age, breaks = 100, prob = TRUE, ann = FALSE, col = "chartreuse")
87  title(xlab = "Age", ylab = "Histogram/Density", "Figure 7: Employee Age - Histogram vs. Normal PDF", cex = 1.2)
88  x <- hist$mids
89  curve(dnorm(x, mean = mu, sd = sigma), add = TRUE, col = "red3", lwd = 2.5)
90
91  #FIGURE 8: Q-Q PLOT - EMPLOYEE AGE
92  qqnorm(Age, ann = FALSE, col = "seagreen")
93  qqline(Age, col = "purple3", lwd = 2.5)
94  title(xlab = "Theoretical Quantiles", ylab = "Sample Quantiles", main = "Figure 8: Q-Q Plot - Employee Age", cex = 1.2)
95
96  #Descriptive Statistics for Age to confirm Skewness and Kurtosis
97  mean(Age);median(Age); skewness(Age);kurtosis(Age)
98
99  #FIGURE 9: SCATTERPLOT - AGE V MONTHLY INCOME
00  plot(Age, MonthlyIncome, xlab = "Employee Age", ylab = "Monthly Income",
01       main = "Figure 9: Scatterplot - Monthly Income based on Age", col = 'red', cex.main = 1.2)
02  abline(lm(UData$MonthlyIncome~UData$Age), lwd = 2.5)
03  cor(Age, MonthlyIncome)
```

Through conducting exploratory data analysis, our consulting specialists have designed two comprehensive models which may allow Globex in anticipating employee attrition. ***Model 1*** utilises a logistic regression containing 15 predictors to determine the probability of staff turnover as a dependent variable. The model utilises a multi-linear model to predict the log-odds of attrition based on the data collected for the chosen variables, for each individual employee. Figure 1.1 demonstrates the use of the log-odds in predicting the probability of attrition. When the model predicts that the employee whose data is used did not leave the firm. However when , the model predicts that the employee did leave the company.

In choosing *Model 1*, the first step was to conduct a logistic regression using all 25 independent variables in proving the log-odds for Attrition. Then, we rationalised this model to 13 variables (*BusinessTravel* and *MaritalStatus* being categorical variables leading to 15 predictors) by using the Bayesian Information Criterion () method. This allowed us to extract only the significant variables from our initial model which disprove the null hypothesis:

Model 2 utilises a classification tree in predicting whether employees stayed or left Globex within the last year. The model uses an intuitive set of rules which can be used by the

company in forecasting staff turnover based on the information collected.

The parameters for *Model 2* were selected using the same 13 variables chosen in *Model 1* as determined through the BIC method. The same variables were chosen in order to maintain consistency within our findings and recommendations.