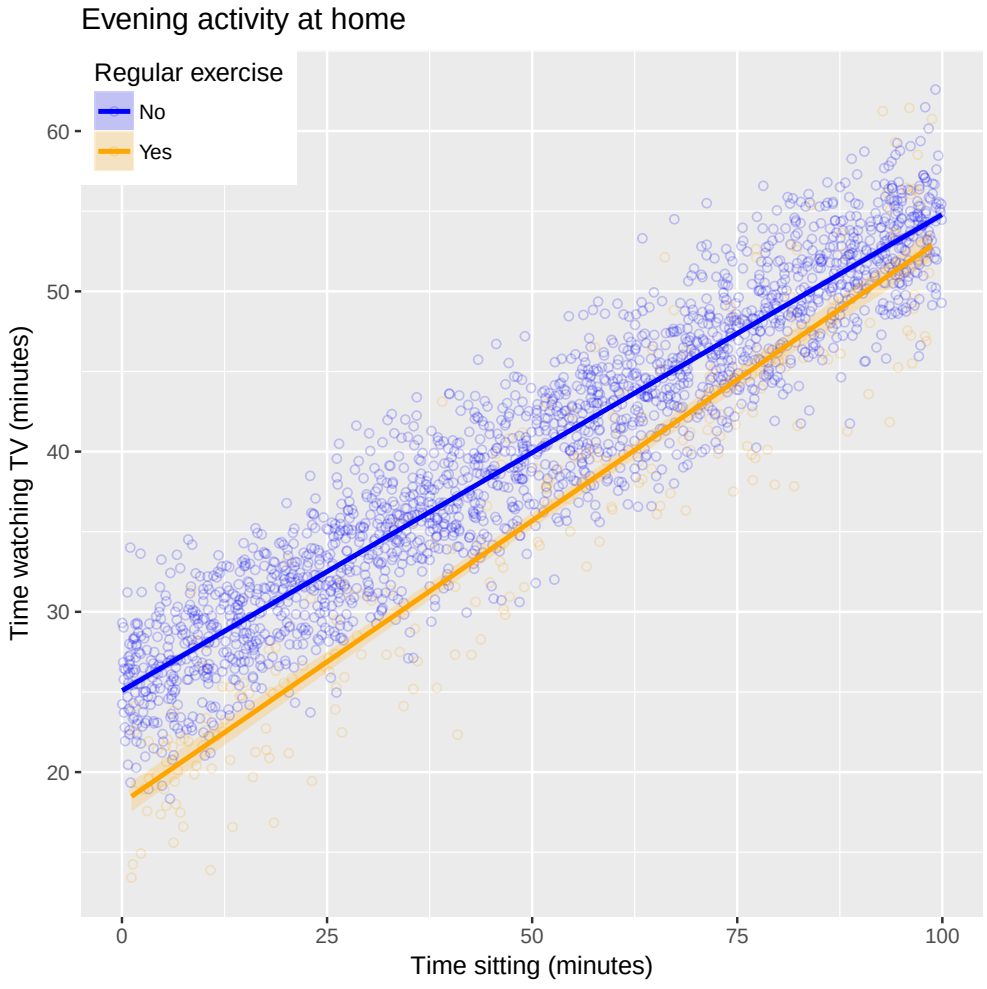
Statistics 3080
Homework 5
David Smith

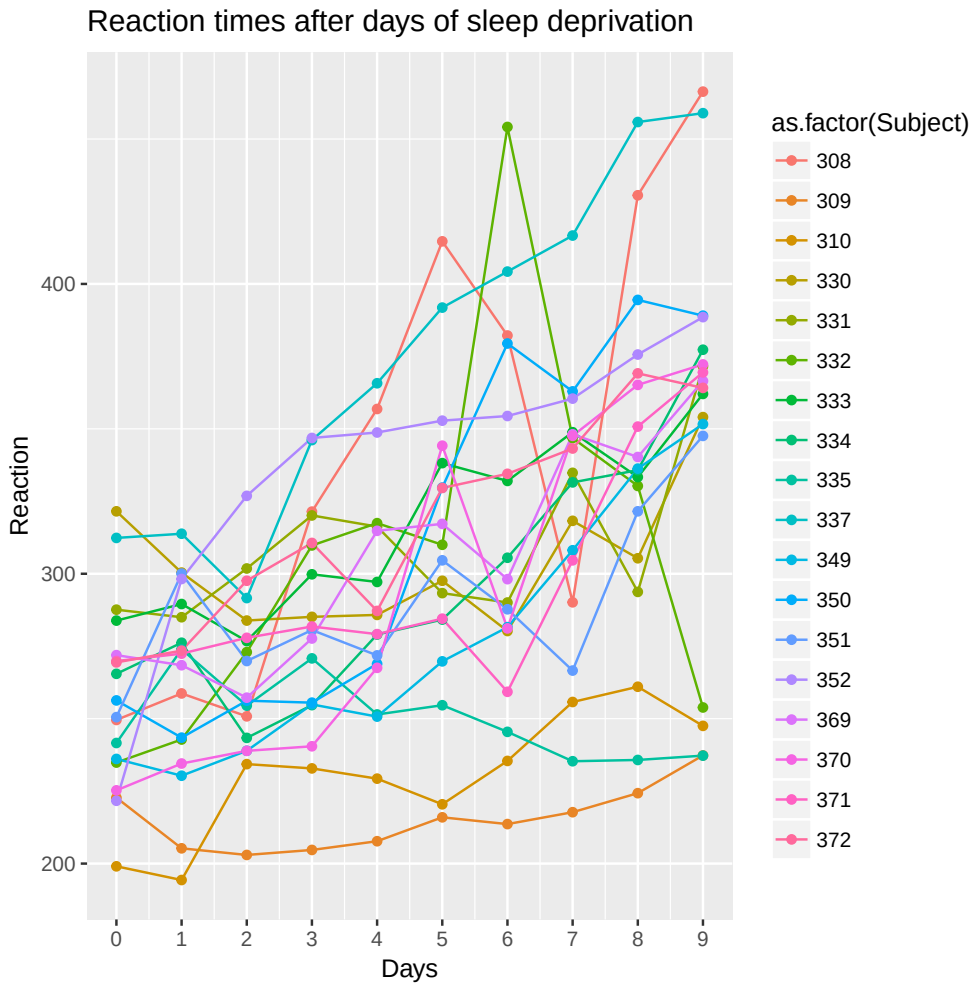
Problem 1a

```
> library(ggplot2)
> tv <- read.table("tv.txt", header=TRUE)
> ggplot(tv, aes(x=x, y=y, colour=as.factor(z), fill=as.factor(z))) +
+   geom_point(shape=21, fill=NA, alpha=0.2) +
+   scale_colour_manual(values=c("1"="blue", "2"="orange"),
+                       labels=c("No", "Yes")) +
+   scale_fill_manual(values=c("1"="blue", "2"="orange"),
+                     labels=c("No", "Yes")) +
+   geom_smooth(method=lm, alpha=0.2) +
+   labs(title="Evening activity at home", x="Time sitting (minutes)",
+        y="Time watching TV (minutes)", colour="Regular exercise",
+        fill="Regular exercise") +
+   theme(legend.position=c(0,1), legend.justification=c(0,1))
```



Problem 1b

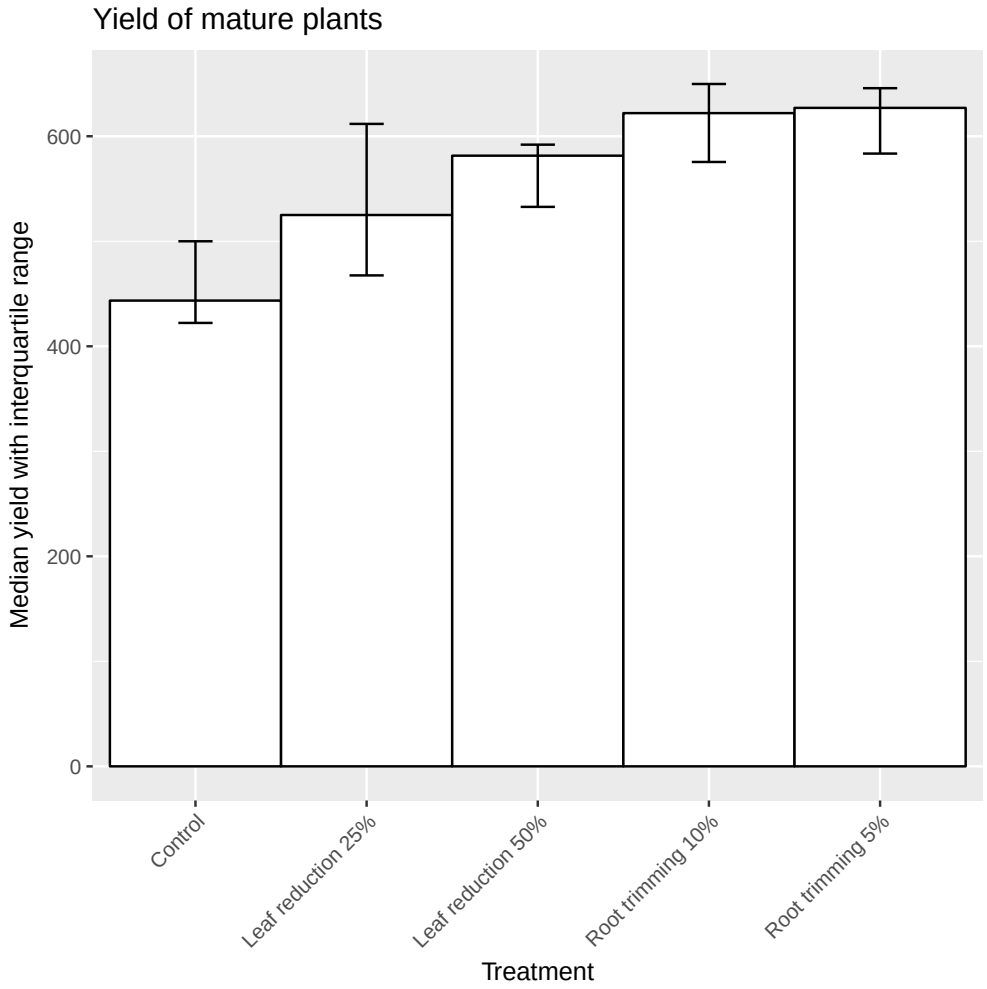
```
> sleep_data <- read.table("sleep.txt", header=TRUE)
> ggplot(sleep_data, aes(x=Days, y=Reaction)) +
+   geom_line(aes(colour=as.factor(Subject))) +
+   geom_point(aes(colour=as.factor(Subject))) +
+   labs(title="Reaction times after days of sleep deprivation") +
+   scale_x_continuous(breaks=0:9)
```



Problem 1c

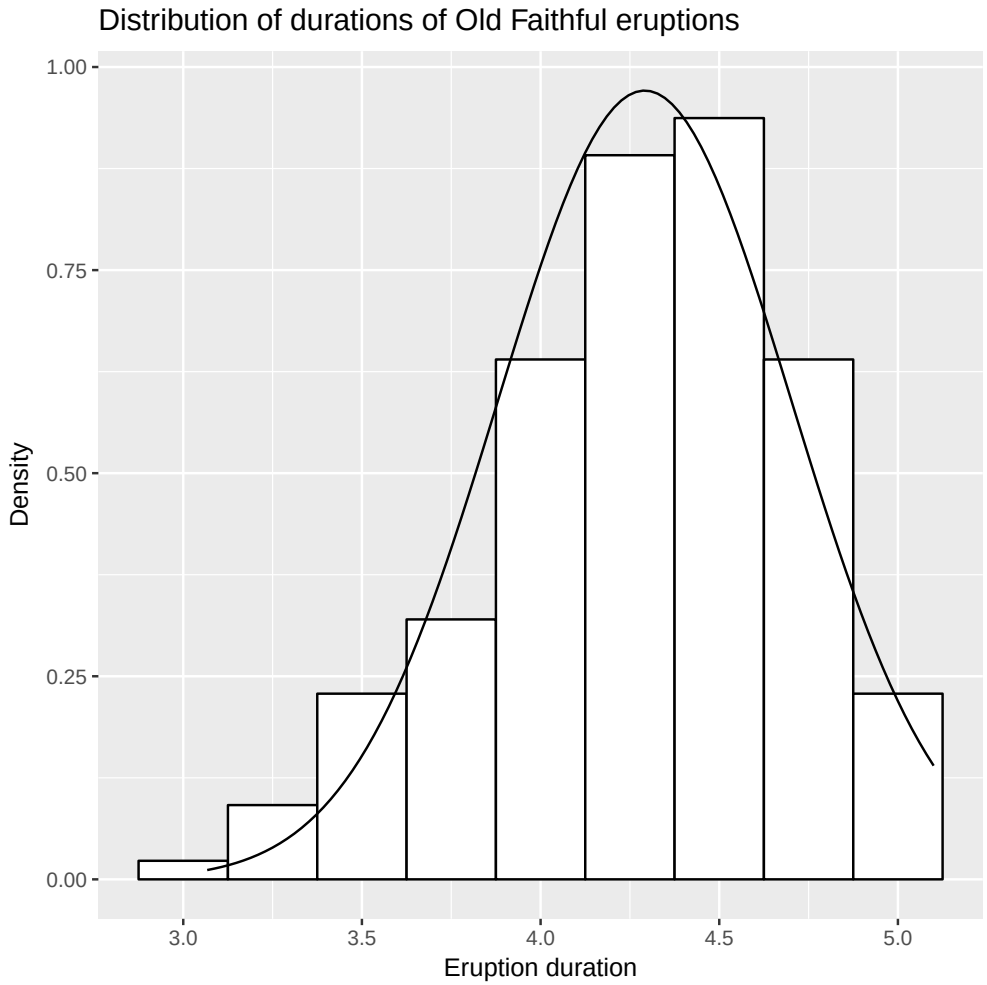
```
> compexp <- read.table("compexp.txt", header=TRUE)
> ggplot(compexp, aes(x=clipping, y=yield)) +
+   stat_summary(fun.y=median, geom="bar", fill="white", colour="black", width=1) +
+   stat_summary(fun.data=median_hilow, fun.args=(conf.int=0.5), geom="errorbar",
+                width=0.2) +
+   labs(title="Yield of mature plants", x="Treatment",
+        y="Median yield with interquartile range") +
+   scale_x_discrete(labels=c("Control", "Leaf reduction 25%", "Leaf reduction 50%",
```

```
+                                "Root trimming 10%", "Root trimming 5%")) +
+   theme(axis.text.x=element_text(angle=45, hjust=1))
```



Problem 1d

```
> erupt <- read.table("erupt.txt", header=TRUE)
> xbar <- mean(erupt$x)
> std_dev <- sd(erupt$x)
> ggplot(erupt) + geom_histogram(aes(x=x, y=..density..), colour="black",
+                                fill="white", binwidth=0.25) +
+   stat_function(fun=dnorm, args=list(mean=xbar, sd=std_dev)) +
+   labs(title="Distribution of durations of Old Faithful eruptions",
+        x="Eruption duration", y="Density")
```



Problem 2a

```
> pop_mean <- 77
> pop_sd <- 31
> n.15 <- 15
> samp.15 <- rnorm(n.15, mean=pop_mean, sd=pop_sd)
> samp.15

 [1]  43.59155 105.31210 108.65184  87.98198  62.95505  48.06359  84.54526
 [8]  57.87434  84.73757  21.56602  33.32627 100.66546  64.57516  75.99737
[15] 122.99168
```

Problem 2b

```
> n.30 <- 30
> samp.30 <- rnorm(n.30, mean=pop_mean, sd=pop_sd)
> samp.30
```

```
 [1]  74.947458  57.494461  65.322982  66.469807  66.126265 109.678521
 [7]  76.288370 139.621861  58.814000  -3.242633 112.337902  20.335735
[13]  77.415412  40.977300 152.087741  47.465299  93.972826  71.080712
[19] 111.725025  92.105275  11.641183 101.716482  93.530173 117.717129
[25]  55.857040  97.482058 101.157836  68.552110  70.602163  93.348264
```

Problem 2c

```
> n.45 <- 45
> samp.45 <- rnorm(n.45, mean=pop_mean, sd=pop_sd)
> samp.45

 [1]  75.72266  51.01626  46.23973 131.55010  73.92203  99.66076  90.06731
 [8]  83.93059 115.27417  34.23632 108.77107  89.16347  92.21873  63.31913
[15]  66.21827  88.53158  87.52476  78.52787 123.20646  61.44499  94.42583
[22]  89.62058  40.50788  62.13131  91.41625  47.60149 109.91930  72.26027
[29] 141.98792  42.26366 162.21674  72.15621  32.91275  73.98907  49.02168
[36]  60.37934  18.51448  80.07626  62.82017  45.53251 110.93932  82.16930
[43]  27.94147  59.46393 177.68883
```

Problem 2d

```
> x_bar.15 <- mean(samp.15)
> x_bar.30 <- mean(samp.30)
> x_bar.45 <- mean(samp.45)
> z_crit <- abs(qnorm(0.025))
> z.15 <- abs((x_bar.15 - pop_mean) / (pop_sd/sqrt(n.15)))
> z.30 <- abs((x_bar.30 - pop_mean) / (pop_sd/sqrt(n.30)))
> z.45 <- abs((x_bar.45 - pop_mean) / (pop_sd/sqrt(n.45)))
> if(z.15 > z_crit) {
+   print("For n = 15: We reject the null hypothesis that the population mean")
+   print("is equal to 77, since the z value for the hypothesis test lies in")
+   print("the rejection region, and conclude that there is evidence to support")
+   print("the claim that the population mean is not equal to 77.")
+ } else {
+   print("For n = 15: We fail to reject the null hypothesis that the population")
+   print("mean is equal to 77, since the z value for the hypothesis test does not")
+   print("fall in the rejection region, and conclude that there is not enough")
+   print("evidence to support the claim that the population mean is not equal")
+   print("to 77.")
+ }

[1] "For n = 15: We fail to reject the null hypothesis that the population"
[1] "mean is equal to 77, since the z value for the hypothesis test does not"
```

```
[1] "fall in the rejection region, and conclude that there is not enough"
[1] "evidence to support the claim that the population mean is not equal"
[1] "to 77."

> if(z.30 > z_crit) {
+   print("For n = 30: We reject the null hypothesis that the population mean")
+   print("is equal to 77, since the z value for the hypothesis test lies in")
+   print("the rejection region, and conclude that there is evidence to support")
+   print("the claim that the population mean is not equal to 77.")
+ } else {
+   print("For n = 30: We fail to reject the null hypothesis that the population")
+   print("mean is equal to 77, since the z value for the hypothesis test does not")
+   print("fall in the rejection region, and conclude that there is not enough")
+   print("evidence to support the claim that the population mean is not equal")
+   print("to 77.")
+ }

[1] "For n = 30: We fail to reject the null hypothesis that the population"
[1] "mean is equal to 77, since the z value for the hypothesis test does not"
[1] "fall in the rejection region, and conclude that there is not enough"
[1] "evidence to support the claim that the population mean is not equal"
[1] "to 77."

> if(z.45 > z_crit) {
+   print("For n = 45: We reject the null hypothesis that the population mean")
+   print("is equal to 77, since the z value for the hypothesis test lies in")
+   print("the rejection region, and conclude that there is evidence to support")
+   print("the claim that the population mean is not equal to 77.")
+ } else {
+   print("For n = 45: We fail to reject the null hypothesis that the population")
+   print("mean is equal to 77, since the z value for the hypothesis test does not")
+   print("fall in the rejection region, and conclude that there is not enough")
+   print("evidence to support the claim that the population mean is not equal")
+   print("to 77.")
+ }

[1] "For n = 45: We fail to reject the null hypothesis that the population"
[1] "mean is equal to 77, since the z value for the hypothesis test does not"
[1] "fall in the rejection region, and conclude that there is not enough"
[1] "evidence to support the claim that the population mean is not equal"
[1] "to 77."
```

Problem 3a

```
> reject.15 <- rep(0, 10000)
> reject.30 <- rep(0, 10000)
> reject.45 <- rep(0, 10000)
> for (i in 1:10000) {
+   samp.15 <- rnorm(n.15, mean=pop_mean, sd=pop_sd)
+   samp.15
+
+   samp.30 <- rnorm(n.30, mean=pop_mean, sd=pop_sd)
+   samp.30
+
+   samp.45 <- rnorm(n.45, mean=pop_mean, sd=pop_sd)
+   samp.45
+
+   x_bar.15 <- mean(samp.15)
+   x_bar.30 <- mean(samp.30)
+   x_bar.45 <- mean(samp.45)
+
+   z.15 <- abs((x_bar.15 - pop_mean) / (pop_sd/sqrt(n.15)))
+   z.30 <- abs((x_bar.30 - pop_mean) / (pop_sd/sqrt(n.30)))
+   z.45 <- abs((x_bar.45 - pop_mean) / (pop_sd/sqrt(n.45)))
+
+   if(z.15 > z_crit) {reject.15[i] <- 1}
+   if(z.30 > z_crit) {reject.30[i] <- 1}
+   if(z.45 > z_crit) {reject.45[i] <- 1}
+ }
```

Problem 3b

```
> mean(reject.15)

[1] 0.0486

> mean(reject.30)

[1] 0.0524

> mean(reject.45)

[1] 0.0513
```

Problem 3c

```
> print("Theoretically, the proportions should be 0.05, since we know")

[1] "Theoretically, the proportions should be 0.05, since we know"
```

```
> print("that the population mean is actually equal to 77. Thus, the")
[1] "that the population mean is actually equal to 77. Thus, the"
> print("null hypothesis that the mean is equal to 77 will be rejected")
[1] "null hypothesis that the mean is equal to 77 will be rejected"
> print("five percent of the time, since we are using a standard significance")
[1] "five percent of the time, since we are using a standard significance"
> print("level of 0.05.")
[1] "level of 0.05."
>
```

References:

- https://stackoverflow.com/questions/15622001/how-to-display-only-integer-values-on-an-axis-using-ggplot2
- https://stackoverflow.com/questions/1330989/rotating-and-spacing-axis-labels-in-ggplot2