Statistics 3080
Homework 3
David Smith

Problem 1a

```
> nym2002 <- read.table("nym2002.txt", header=TRUE)
```

Problem 1b

```
> length(nym2002[,"time"])

[1] 68
```

Problem 1c

```
> paste("The fastest time is", min(nym2002[,"time"]))

[1] "The fastest time is 163.9333333"

> paste("The slowest time is", max(nym2002[,"time"]))

[1] "The slowest time is 467.1833333"
```

Problem 1d

```
> fast_hours <- floor(min(nym2002[,"time"]) / 60)
> fast_minutes <- min(nym2002[,"time"]) - 60*fast_hours
> slow_hours <- floor(max(nym2002[,"time"]) / 60)
> slow_minutes <- max(nym2002[,"time"]) - 60*slow_hours
> paste("The fastest time is", fast_hours, "hours and", fast_minutes, "minutes.")

[1] "The fastest time is 2 hours and 43.9333333 minutes."

> paste("The slowest time is", slow_hours, "hours and", slow_minutes, "minutes.")

[1] "The slowest time is 7 hours and 47.1833333 minutes."
```

Problem 1e

```
> woman_slow <- max(nym2002[nym2002[,"gender"] == "Female", "time"])
> length(nym2002[nym2002[,"time"] > woman_slow, "time"])

[1] 0
```

Problem 1f

```
> fastest_time <- min(nym2002[,"time"])
> as.character(nym2002[nym2002[,"time"] == fastest_time, "home"])

[1] "GER"
```

Problem 1g

```
> man_slow <- max(nym2002[nym2002[,"gender"] == "Male", "time"])
> nym2002[nym2002[,"time"] == man_slow, "age"]

[1] 67
```

Problem 1h

```
> nym2002[nym2002[,"time"] == fastest_time, "place"]

[1] 200
```

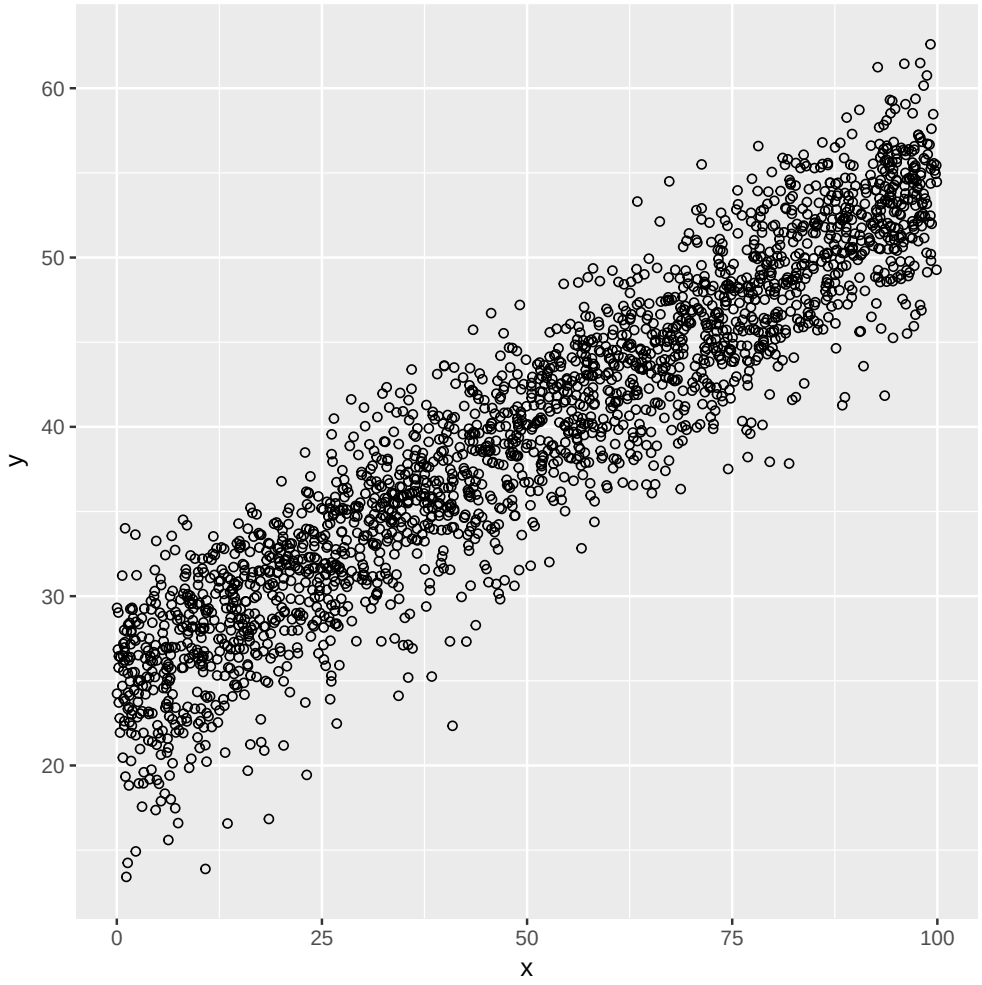
Problem 1i

```
> length(nym2002[nchar(as.character(nym2002[,"home"])) == 3, "home"])

[1] 30
```

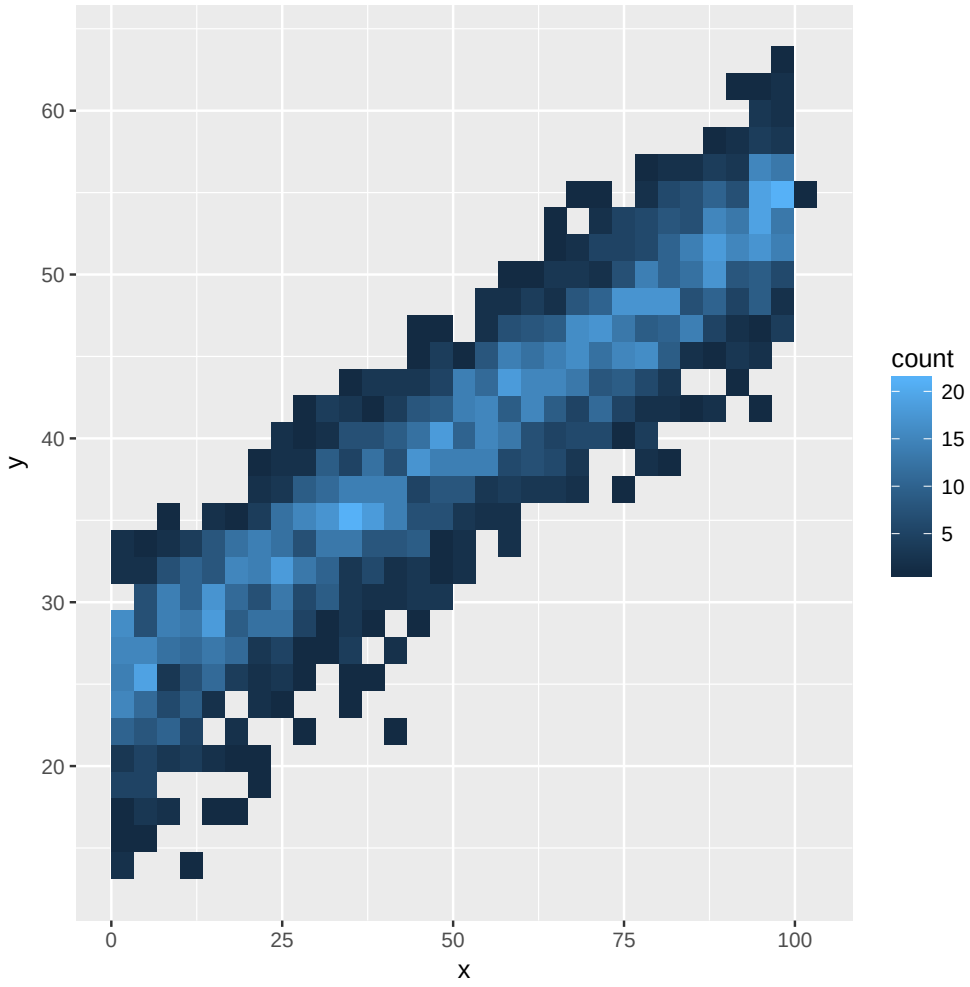
Problem 2a

```
> library(ggplot2)
> tv <- read.table("tv.txt", header=TRUE)
> ggplot(tv, aes(x=x, y=y)) + geom_point(shape=21, fill=NA)
```



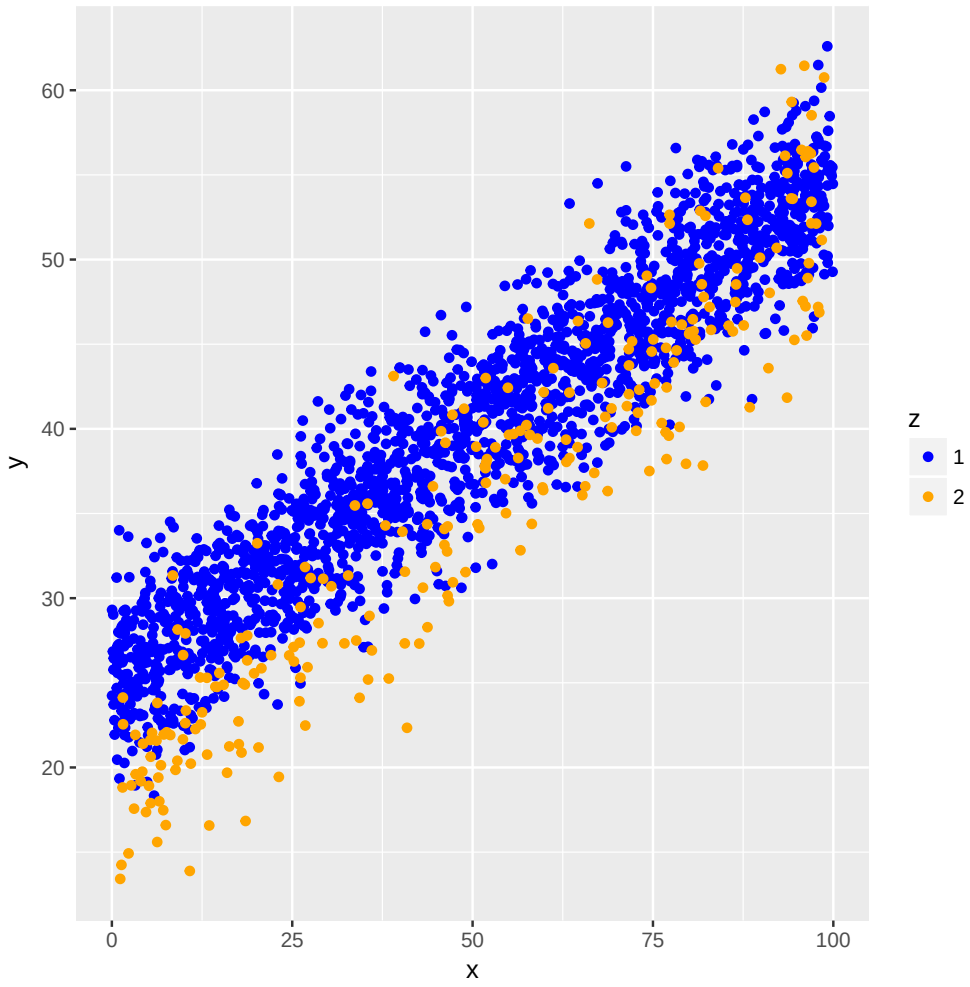
Problem 2b

```
> ggplot(tv, aes(x=x, y=y)) + stat_bin2d()
```



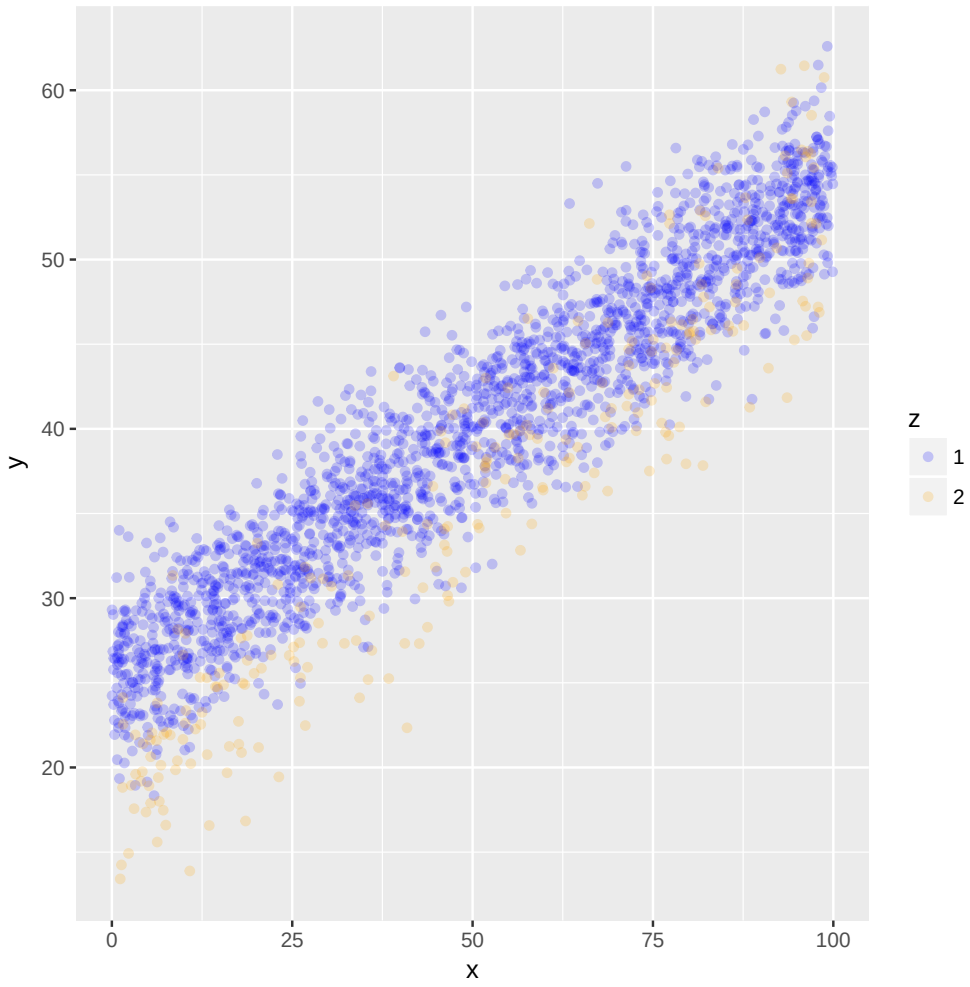
Problem 2c

```
> ggplot(tv, aes(x=x, y=y, colour=as.factor(z))) + geom_point() +
+   scale_colour_manual("z", values=c("1"="blue", "2"="orange"))
```



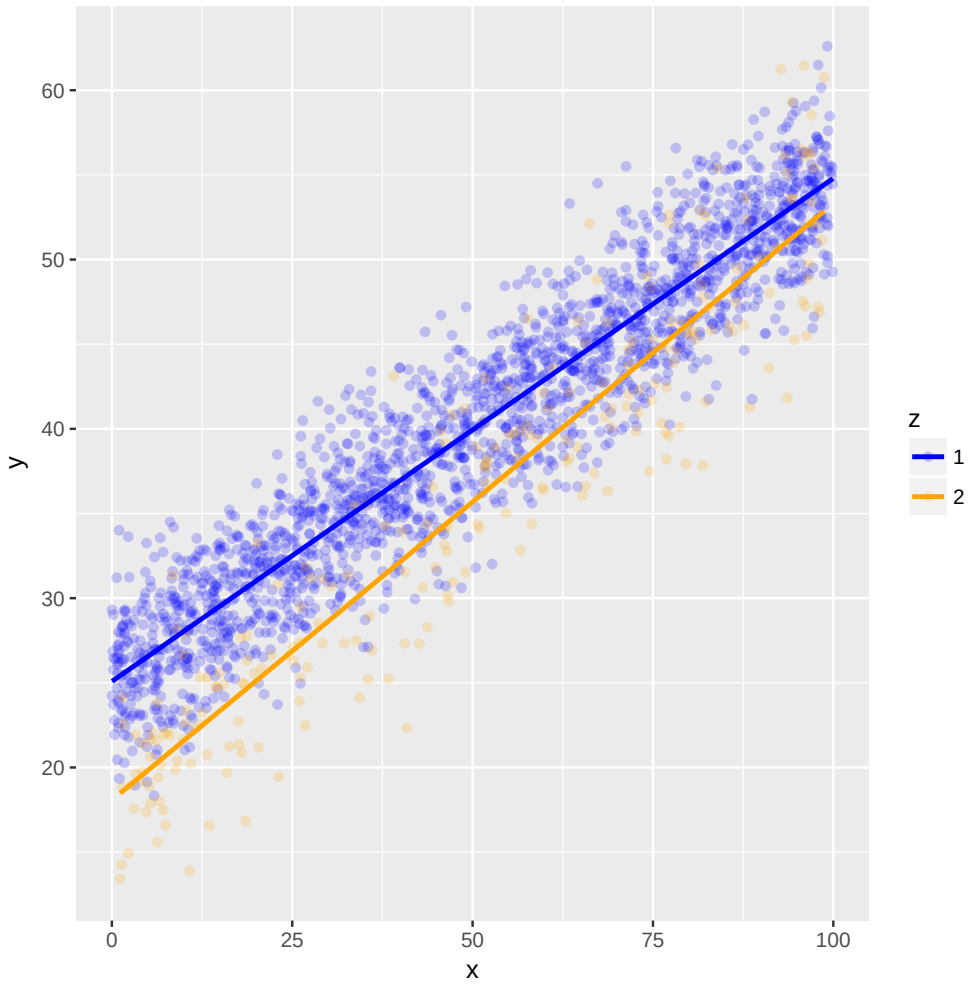
Problem 2d

```
> ggplot(tv, aes(x=x, y=y, colour=as.factor(z))) + geom_point(alpha=0.2) +
+   scale_colour_manual("z", values=c("1"="blue", "2"="orange"))
```



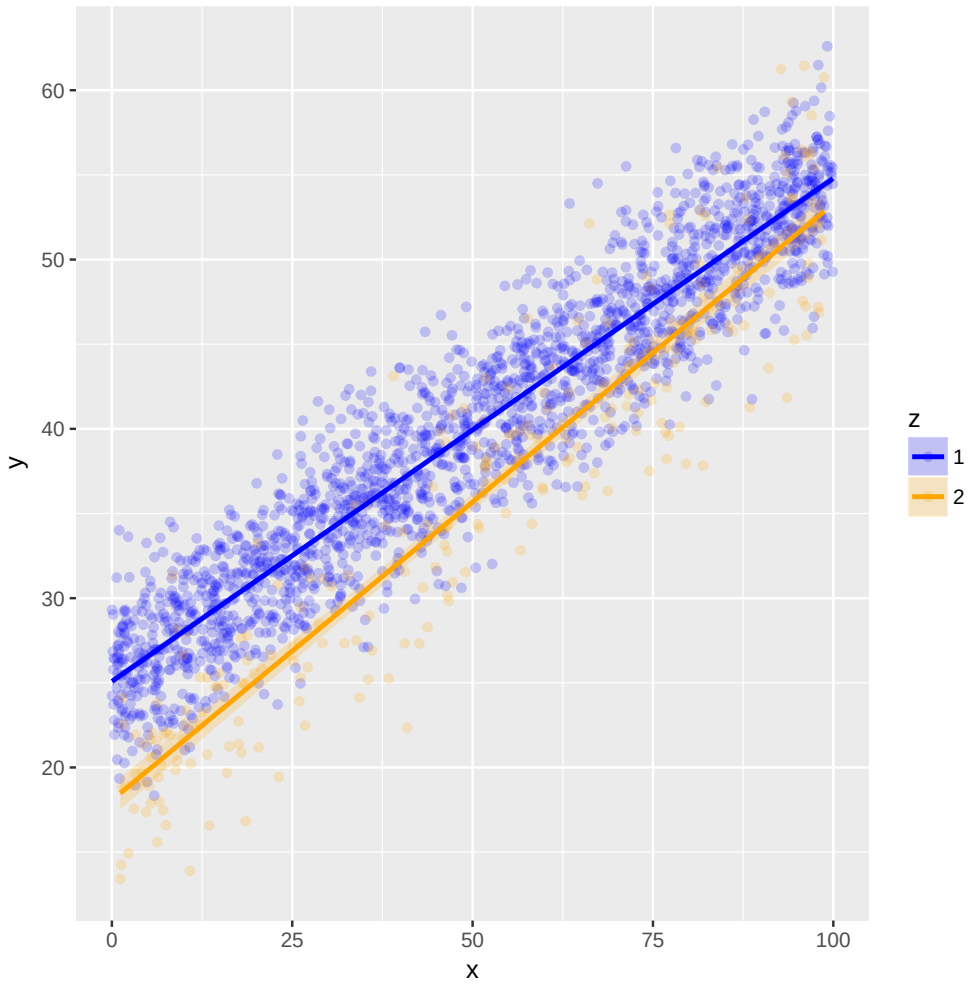
Problem 2e

```
> ggplot(tv, aes(x=x, y=y, colour=as.factor(z))) + geom_point(alpha=0.2) +
+   scale_colour_manual("z", values=c("1"="blue", "2"="orange")) +
+   geom_smooth(method=lm, se=FALSE)
```



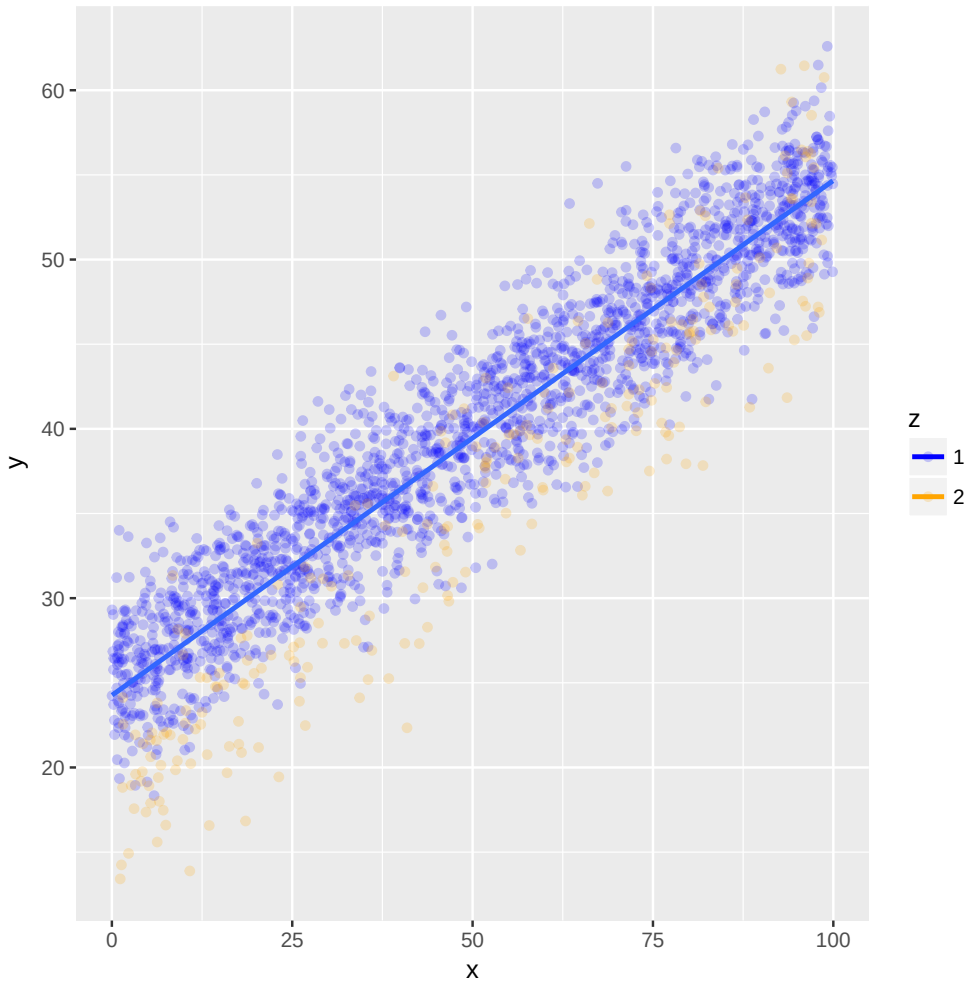
Problem 2f

```
> ggplot(tv, aes(x=x, y=y, colour=as.factor(z), fill=as.factor(z))) +
+   geom_point(alpha=0.2) +
+   scale_colour_manual("z", values=c("1"="blue", "2"="orange")) +
+   scale_fill_manual("z", values=c("1"="blue", "2"="orange")) +
+   geom_smooth(method=lm, alpha=0.2)
```



Problem 2g

```
> ggplot(tv, aes(x=x, y=y, colour=as.factor(z))) + geom_point(alpha=0.2) +
+   scale_colour_manual("z", values=c("1"="blue", "2"="orange")) +
+   geom_smooth(method=lm, aes(group=1), se=FALSE)
```



References:

- https://stackoverflow.com/questions/11134812/how-to-find-the-length-of-a-string-in-r
- https://stackoverflow.com/questions/15726907/ggplot-scatterplot-points-with-no-fill
- https://stackoverflow.com/questions/47677176/how-to-add-the-confidence-interval-in-scatter-plot-ggplot-in-r
- https://stackoverflow.com/questions/37348719/ggplot2-single-regression-line-when-colour-is-coded-for-by-a-variable