

Class 05 Data Visualization

Nadia Haghani

2021-10-12

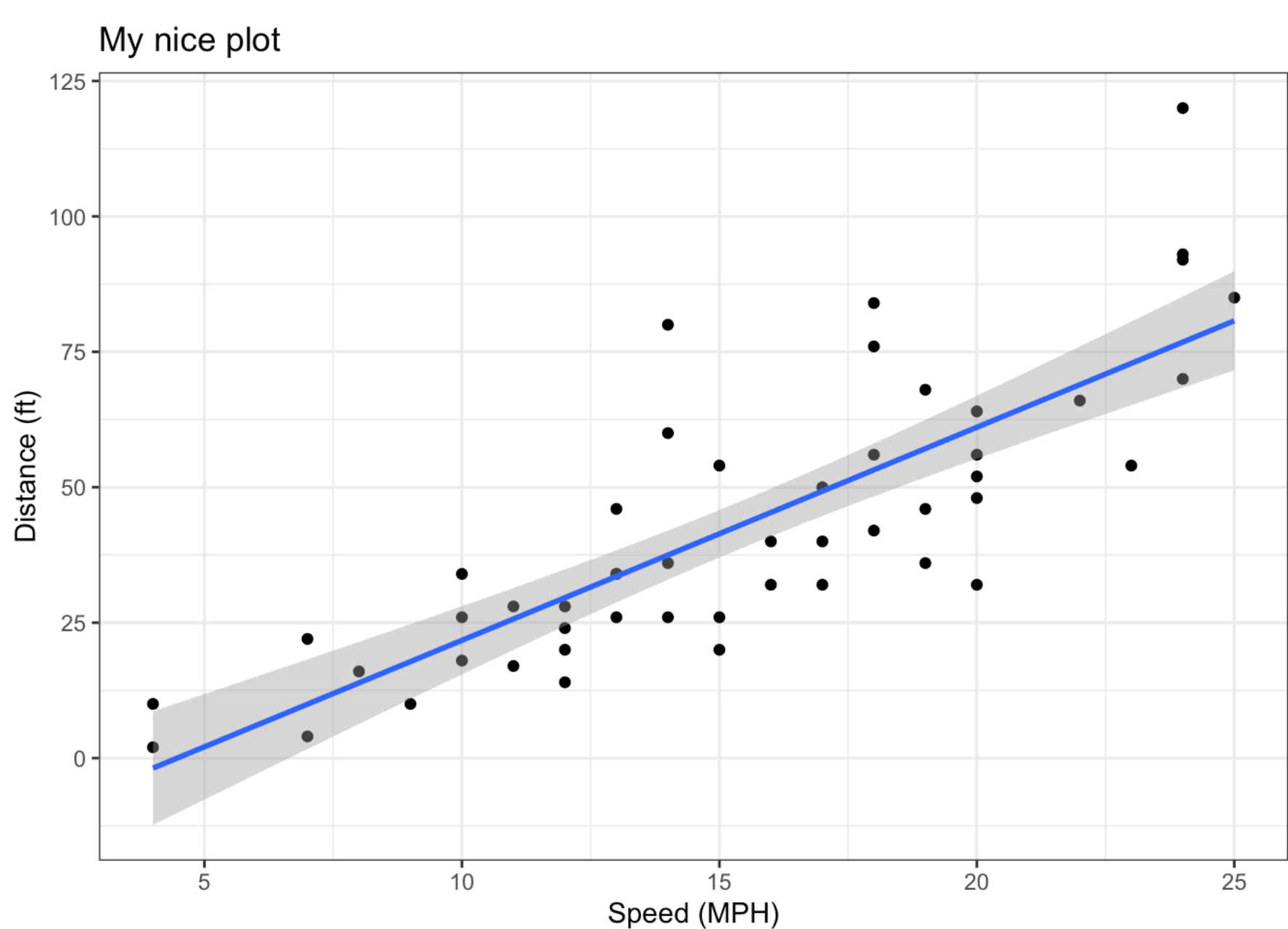
```
# Lets start with a scatterplot

library(ggplot2) #load before using

# Every ggplot has a data + aes + geoms
p <- ggplot(data = cars) +
  aes(x=speed, y=dist) +
  geom_point() +
  geom_smooth(method="lm")

p + labs(title = "My nice plot",
         x= "Speed (MPH)",
         y = "Distance (ft)") +
  theme_bw()
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
# A More complicated dataset

# First read the dataset
url <- "https://bioboot.github.io/bimml43_S20/class-material/up_down_expression.txt"
genes <- read.delim(url)
head(genes)
```

##	Gene	Condition1	Condition2	State
## 1	A4GNT	-3.6808610	-3.4401355	unchanging
## 2	AAAS	4.5479580	4.3864126	unchanging
## 3	AASDH	3.7190695	3.4787276	unchanging
## 4	AATF	5.0784720	5.0151916	unchanging
## 5	AATK	0.4711421	0.5598642	unchanging
## 6	AB015752.4	-3.6808610	-3.5921390	unchanging

```
# How many genes in dataset?
nrow(genes)
```

```
## [1] 5196
```

```
# How to access State col
table(genes$State)
```

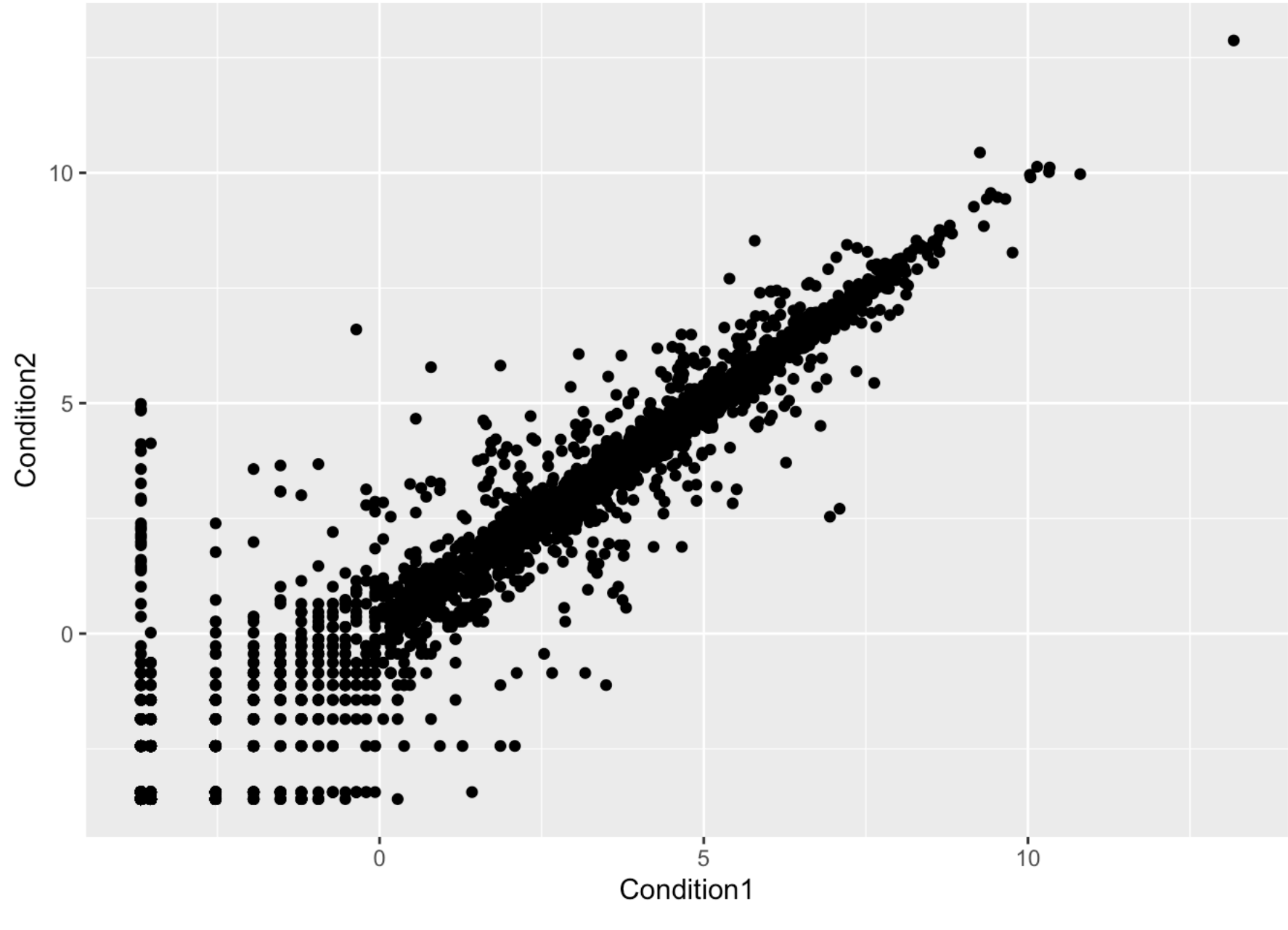
##	down	unchanging	up
##	72	4997	127

```
# What % are up/down
prec <- table(genes$State) / nrow(genes) * 100
round(prec, 2)
```

##	down	unchanging	up
##	1.39	96.17	2.44

```
# Make a basic scatterplot of genes dataset

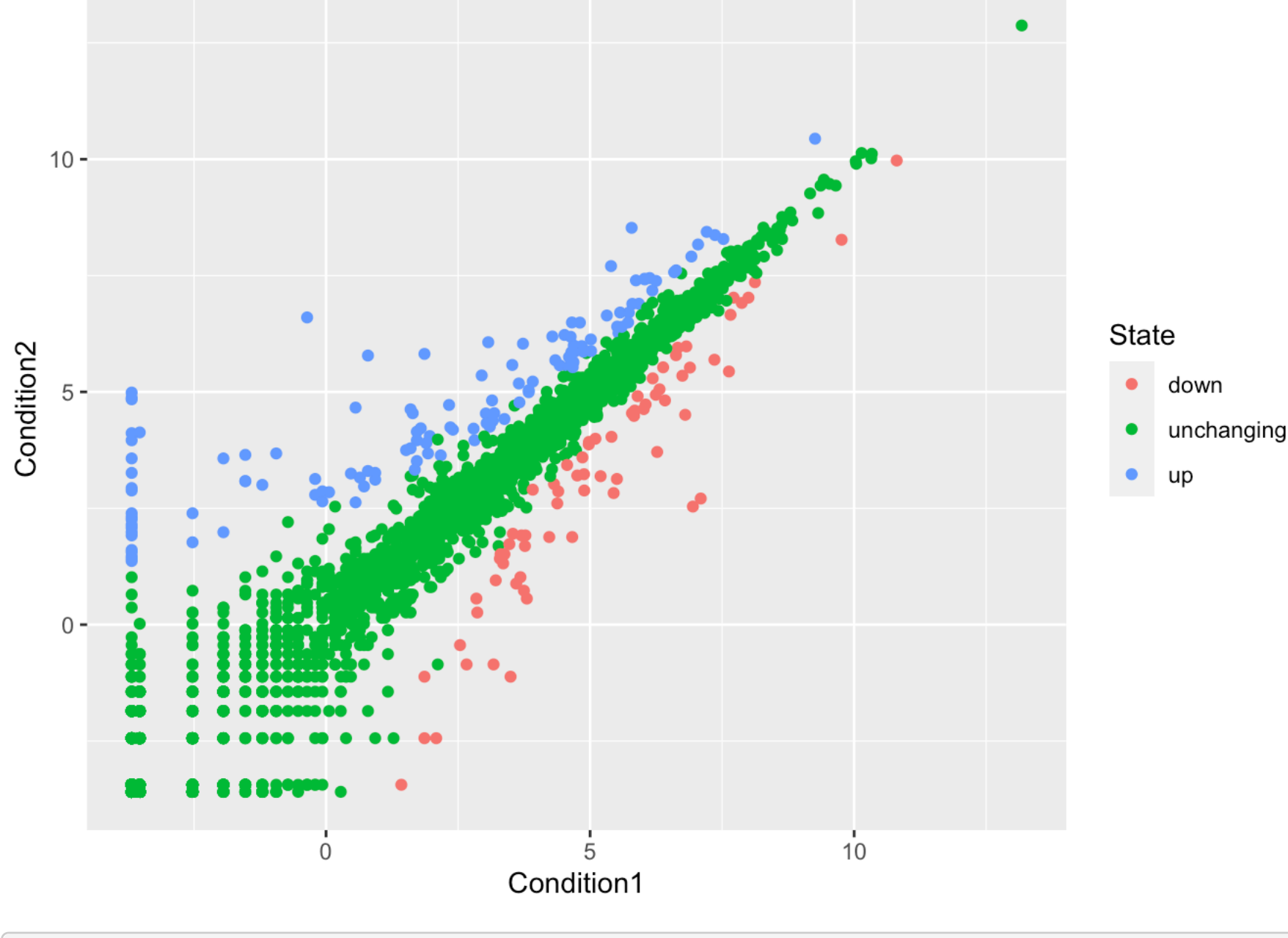
ggplot(data = genes) +
  aes(x = Condition1, y = Condition2) +
  geom_point()
```



```
# Adding color to the state of the gene

p <- ggplot(genes) +
  aes(x=Condition1, y=Condition2, col=State) +
  geom_point()

p
```



```
# Change color

p + scale_colour_manual(values= c("blue", "grey", "red")) +
  labs(title = "Gene Expression Changes Upon Drug Treatment",
       x = "Control (no drug)",
       y = "Drug Treatment")
```

