

Class05.R

stefaniehodapp

2021-10-13

```
# Class 05: data visualization
```

```
# Using the ggplot2 package
```

```
# Load package  
library(ggplot2)
```

```
# We will see this inbuilt "cars" dataset first  
head(cars)
```

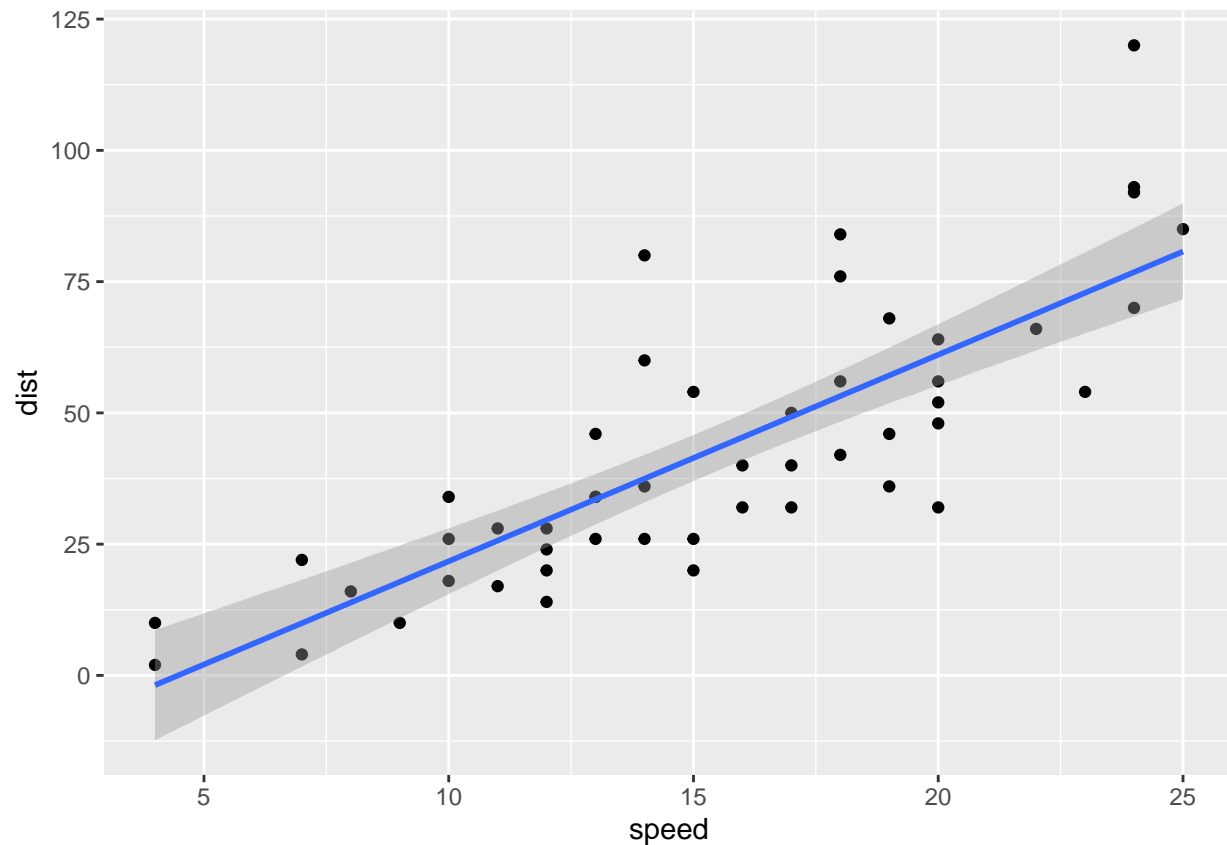
```
##   speed dist  
## 1     4    2  
## 2     4   10  
## 3     7    4  
## 4     7   22  
## 5     8   16  
## 6     9   10
```

```
# All ggplots have at least 3 layers:
```

```
#   data + aes + geoms
```

```
ggplot(data=cars) + aes(x=speed, y=dist) + geom_point() +  
  geom_smooth(method="lm")
```

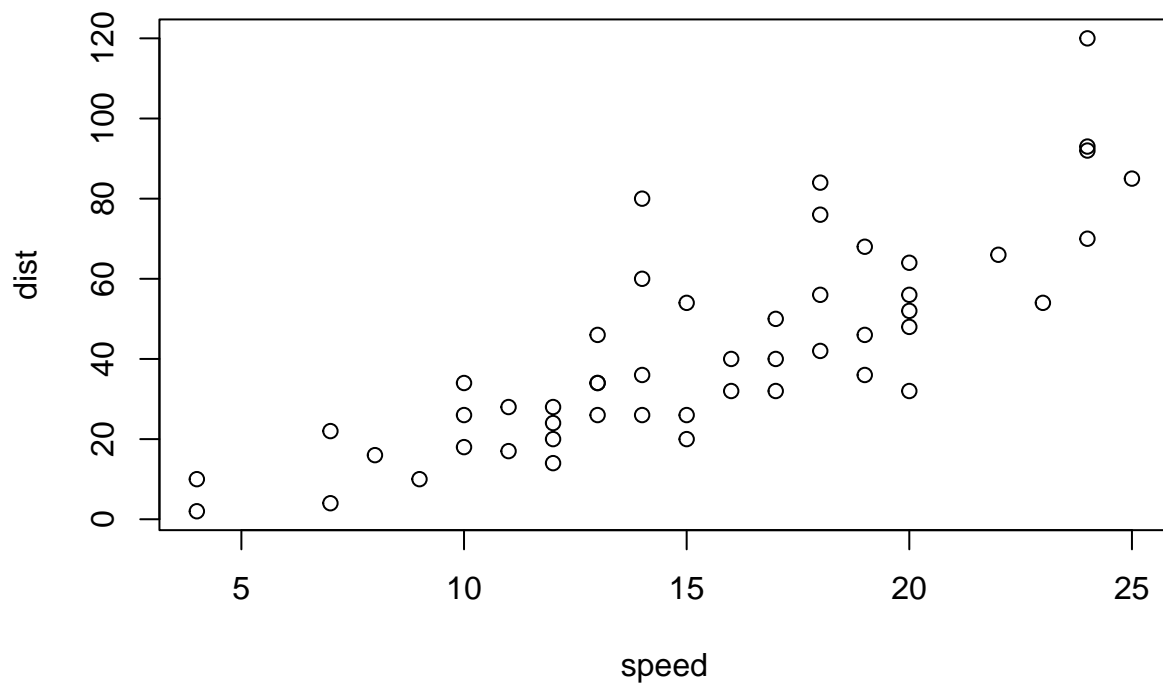
```
## 'geom_smooth()' using formula 'y ~ x'
```



```
labs(title="Stopping Distance of Old Cars",
      x="Speed (MPH)", y="Stopping Sistance (ft)")
```

```
## $x
## [1] "Speed (MPH)"
##
## $y
## [1] "Stopping Sistance (ft)"
##
## $title
## [1] "Stopping Distance of Old Cars"
##
## attr("class")
## [1] "labels"
```

```
# Side note: ggplot is not the only graphics system
# A very popular one is "base" R graphics
plot(cars)
```



```
# Lab report questions
```

```
url <- "https://bioboot.github.io/bimm143_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
```

```
nrow(genes) # number of rows
```

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

```
colnames(genes) # column names
```

```
## [1] "Gene"      "Condition1" "Condition2" "State"
```

```
ncol(genes) # number of columns
```

```
## [1] 4
```

```
table(genes$State) # number of genes per state
```

```
##
##      down  unchanged      up
##      72      4997      127
```

```
table(genes$State)/nrow(genes) * 100 # fraction of total up-regulated genes
```

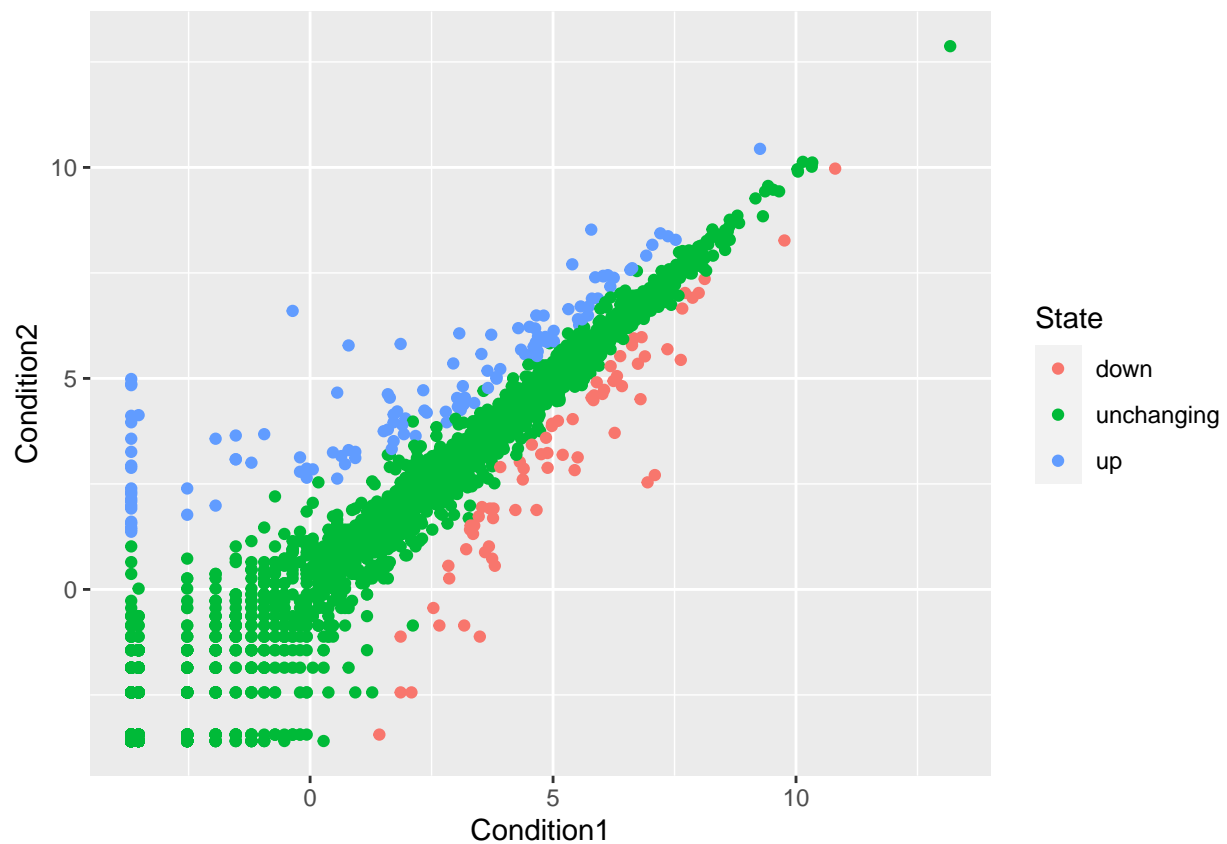
```
##
##      down  unchanged      up
##  1.385681  96.170131  2.444188
```

```
round(table(genes$State)/nrow(genes) * 100, 3) # 3 significant figures
```

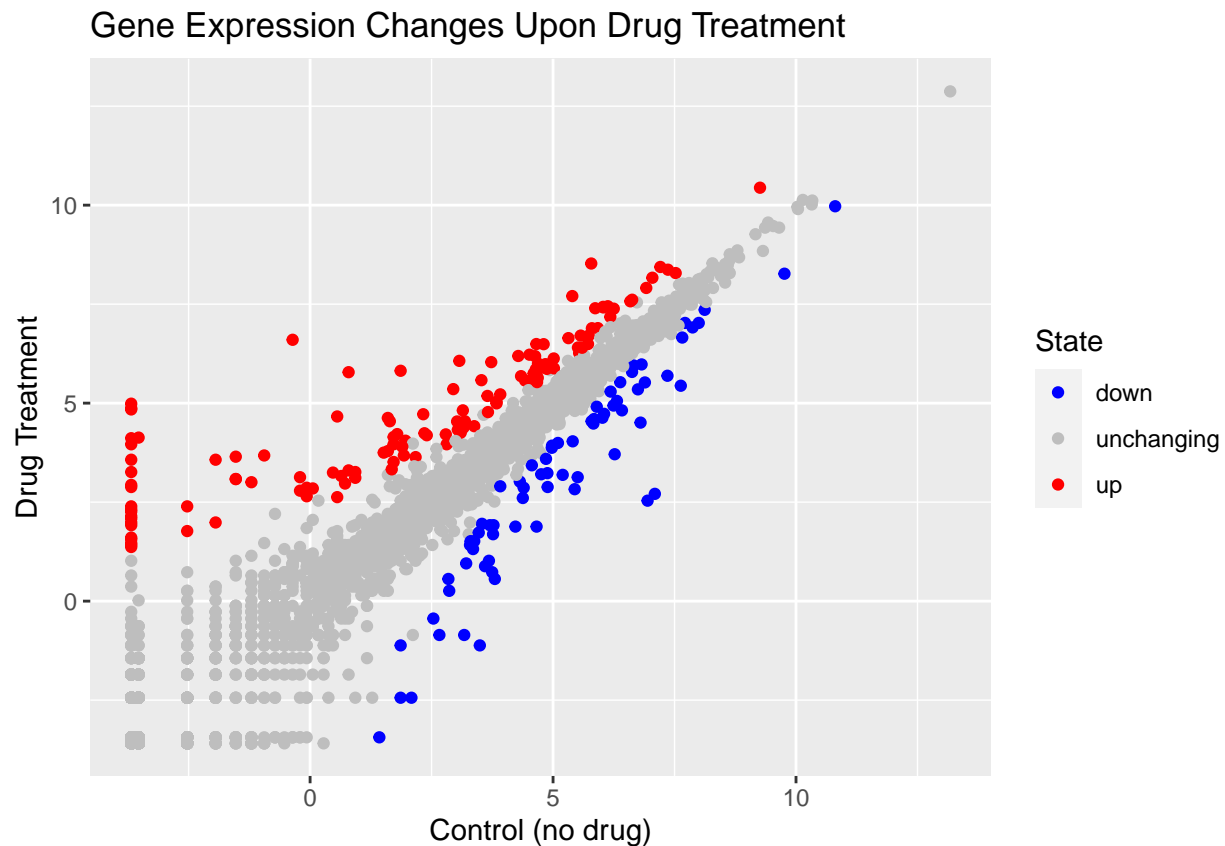
```
##
##      down  unchanged      up
##    1.386    96.170    2.444
```

```
# make scatter plot
```

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



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



```
# Install gapminder dataset
# Contains economic and demographic data about various countries since 1952
# install.packages("gapminder")
library(gapminder)

# read-in data from online
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv"

gapminder <- read.delim(url)

# install.packages("dplyr")
library(dplyr)

##
## Attaching package: 'dplyr'

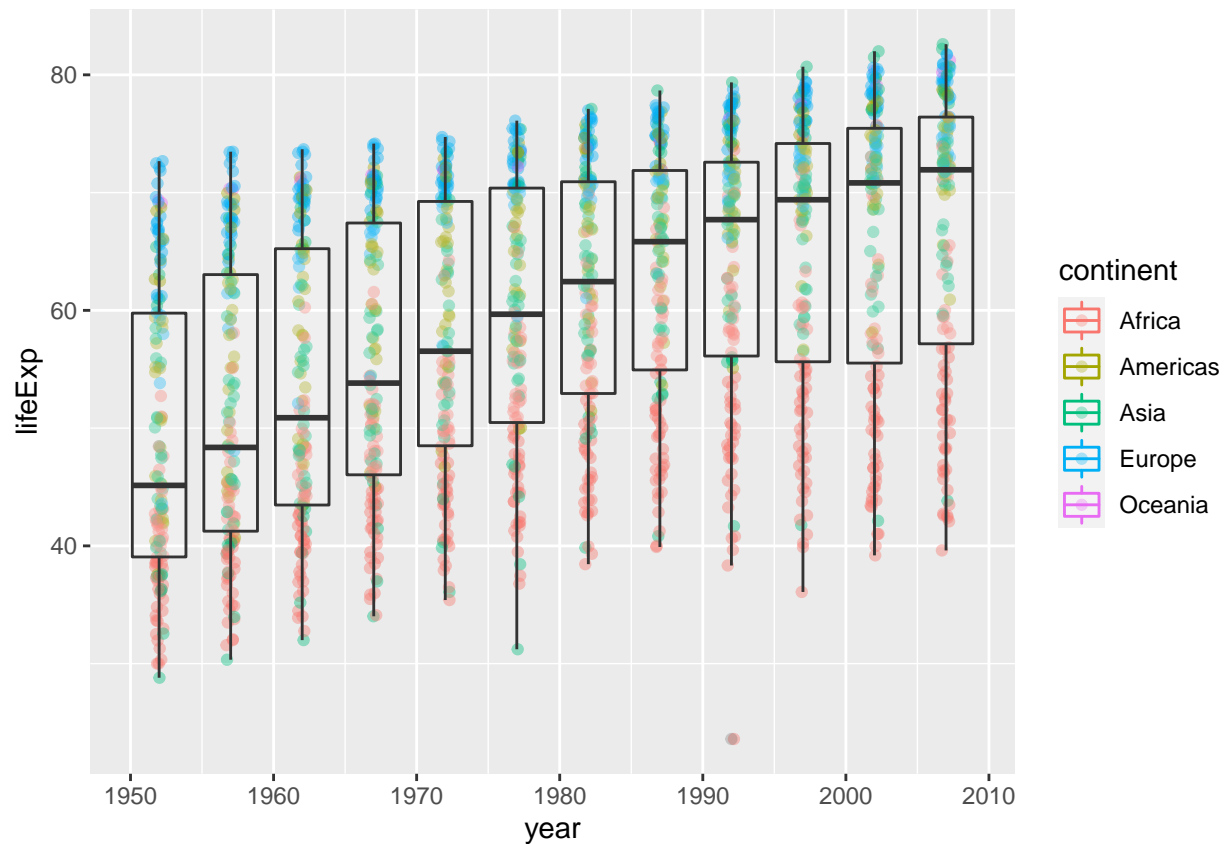
## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
```

```
##
##   intersect, setdiff, setequal, union
```

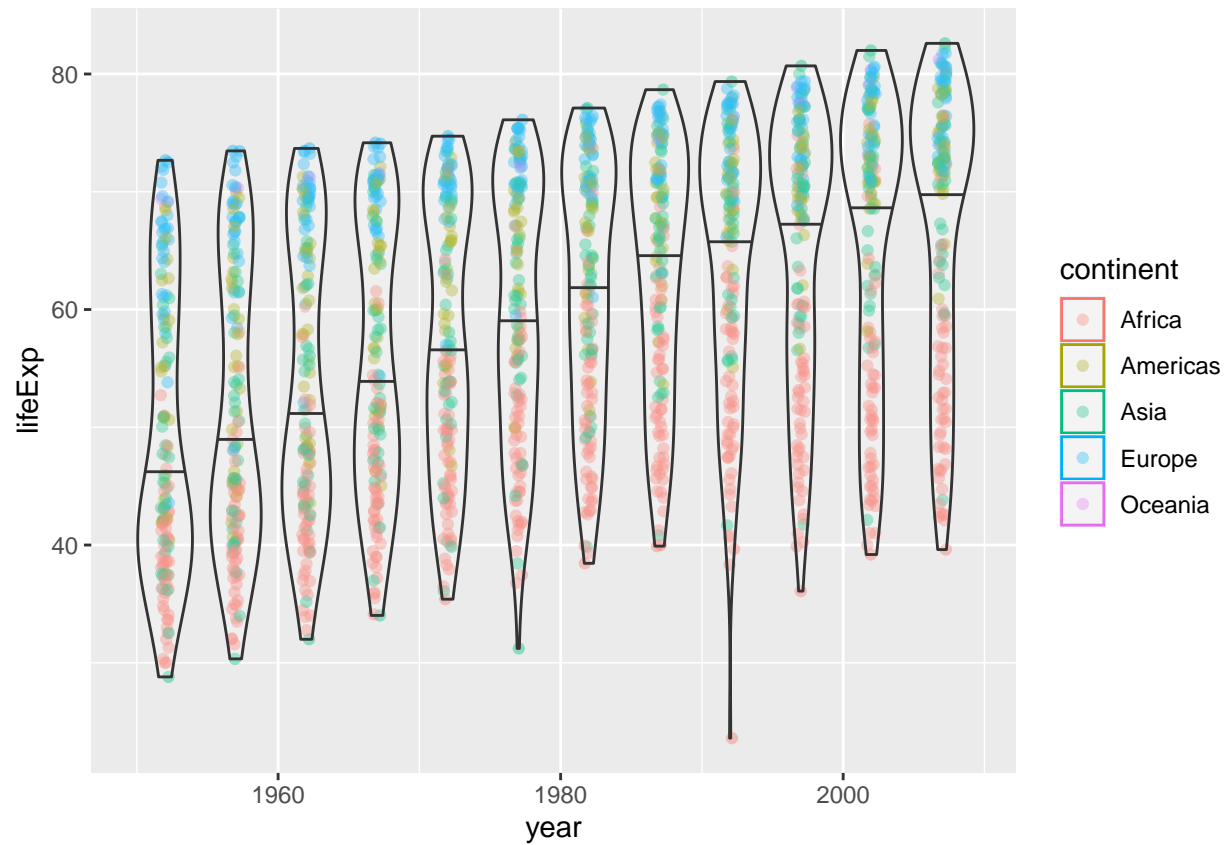
```
# Let's make a jitter plot of year vs lifeExp
```

```
ggplot(gapminder) + aes(x=year, y=lifeExp, col=continent) +
  geom_jitter(width=0.3,alpha=0.4) + geom_boxplot( aes(group=year), alpha=0.2 )
```



```
# Violin plot
```

```
lifeExp_by_continent <- ggplot(gapminder) + aes(x=year, y=lifeExp, col=continent) +
  geom_jitter(width=0.3,alpha=0.4) + geom_violin(aes(group=year), alpha=0.2, draw_quantiles=0.5)
lifeExp_by_continent
```



```
# Install the plotly
# install.packages("plotly")
# library(plotly)
# ggplotly()

# Filter by the year 2007
gapminder_2007 <- gapminder %>% filter(year==2007)
```