

Exo supp chapitre 6

Imane Ez-Zammoury

24/03/2021

Setups

```
library(rWSBIM1207)
library(tidyverse)
library(ggplot2)
```

Exercice 1

Analyse the data to answer the question Do men drink more than women according to age and working status? Now reproduce the figure below.

```
data(beers)

mw <- beers %>% filter(!is.na(Consumption), !is.na(Age), !is.na(Gender), !is.na(Work)) %>%
  group_by(Gender, Age, Work) %>% summarise(moyenne_conso = mean(Consumption))

## `summarise()` has grouped output by 'Gender', 'Age'. You can override using the `.groups` argument.
mw

## # A tibble: 16 x 4
## # Groups:   Gender, Age [8]
##   Gender Age Work      moyenne_conso
##   <fct> <int> <fct>          <dbl>
## 1 Female  25 Employed          220
## 2 Female  25 Unemployed        190
## 3 Female  35 Employed          190
## 4 Female  35 Unemployed        143
## 5 Female  45 Employed          175
## 6 Female  45 Unemployed        132.
## 7 Female  55 Employed          128.
## 8 Female  55 Unemployed        132.
## 9 Male    25 Employed          110
## 10 Male   25 Unemployed        162.
## 11 Male   35 Employed          153.
## 12 Male   35 Unemployed        188.
## 13 Male   45 Employed          188.
## 14 Male   45 Unemployed        233.
## 15 Male   55 Employed          210
## 16 Male   55 Unemployed        313.

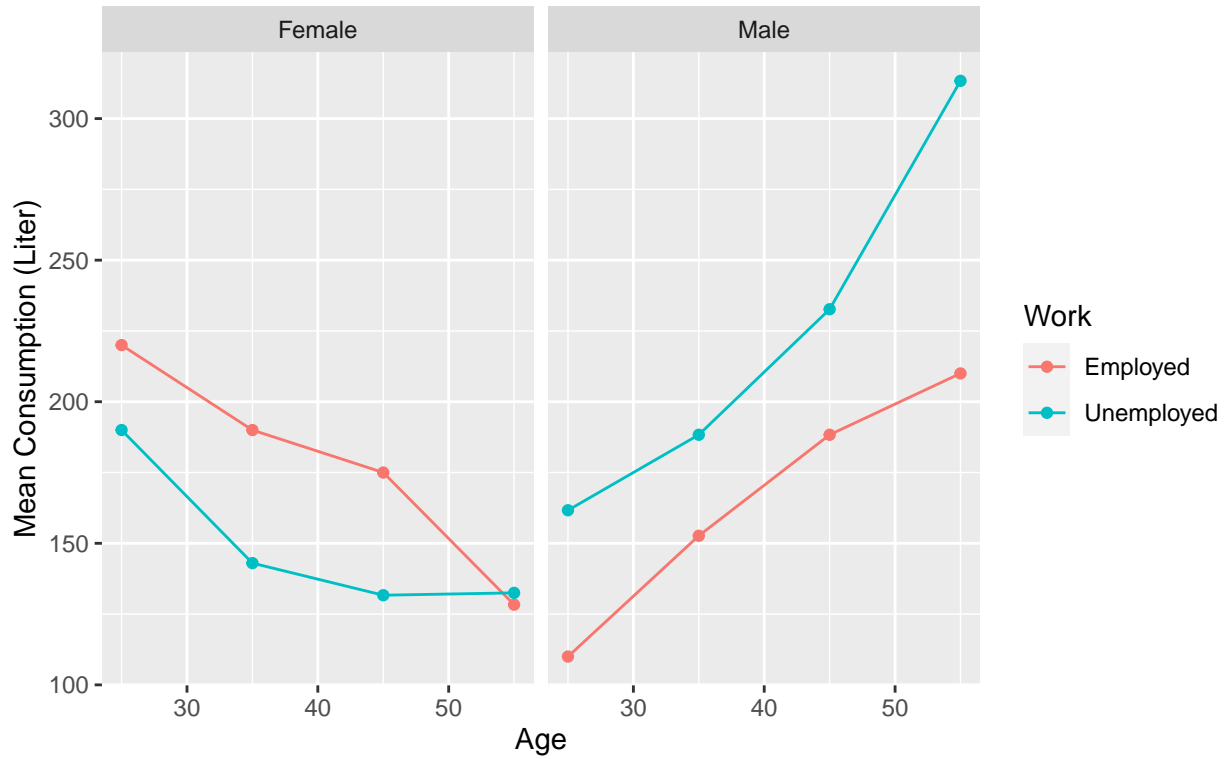
ggplot(mw, aes(x = Age, y = moyenne_conso, colour=Work)) + geom_point() + geom_line() +
  facet_wrap(~Gender) + labs(title = "Means consumption of beers according to age,
```

```

    employment status and gender",
    x = "Age",
    y = "Mean Consumption (Liter)")

```

Means consumption of beers according to age,
employment status and gender



Exercise 2

Question 1

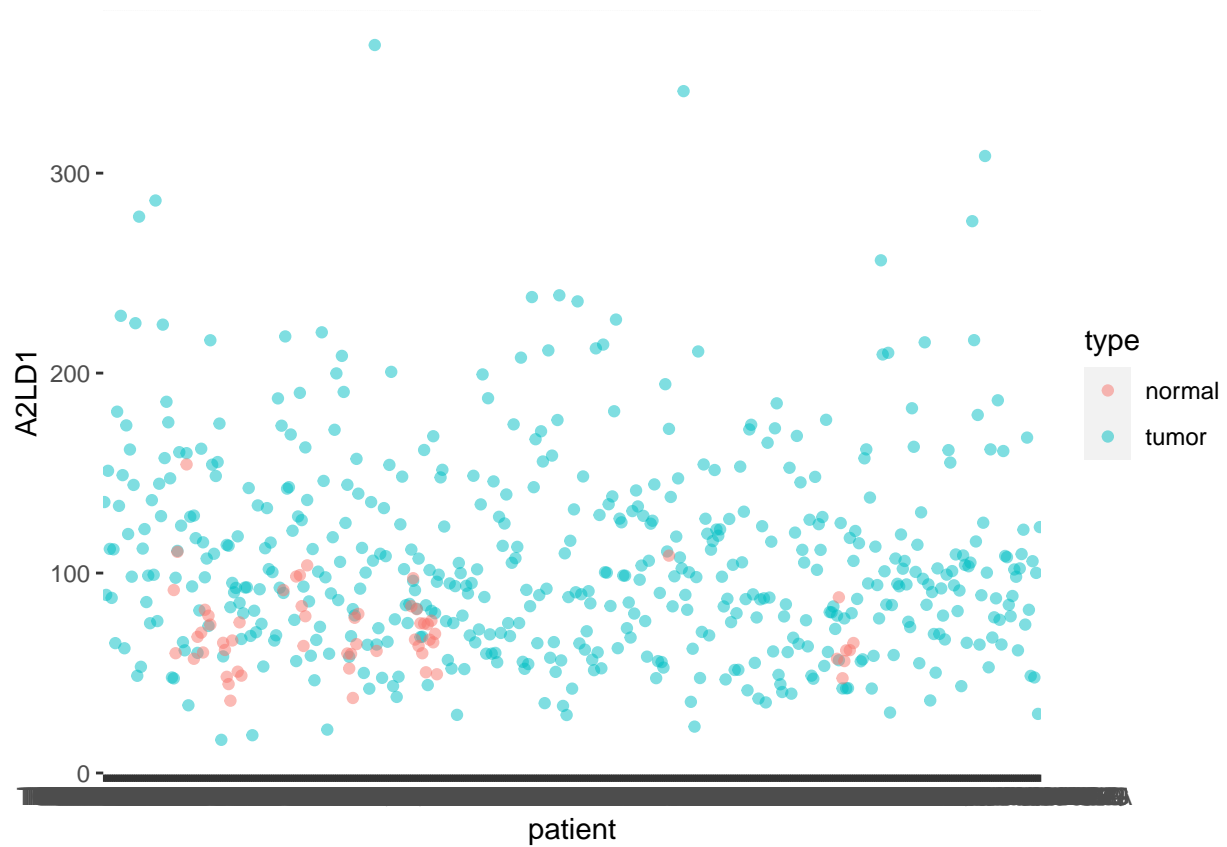
Using `geom_point`, draw a plot showing distribution of expression levels of A2LD1 in normal tissue samples and in primary tumor samples.

```

data("expression")

ggplot(expression, aes(x = patient , y = A2LD1, colour=type)) + geom_point(alpha=0.5)

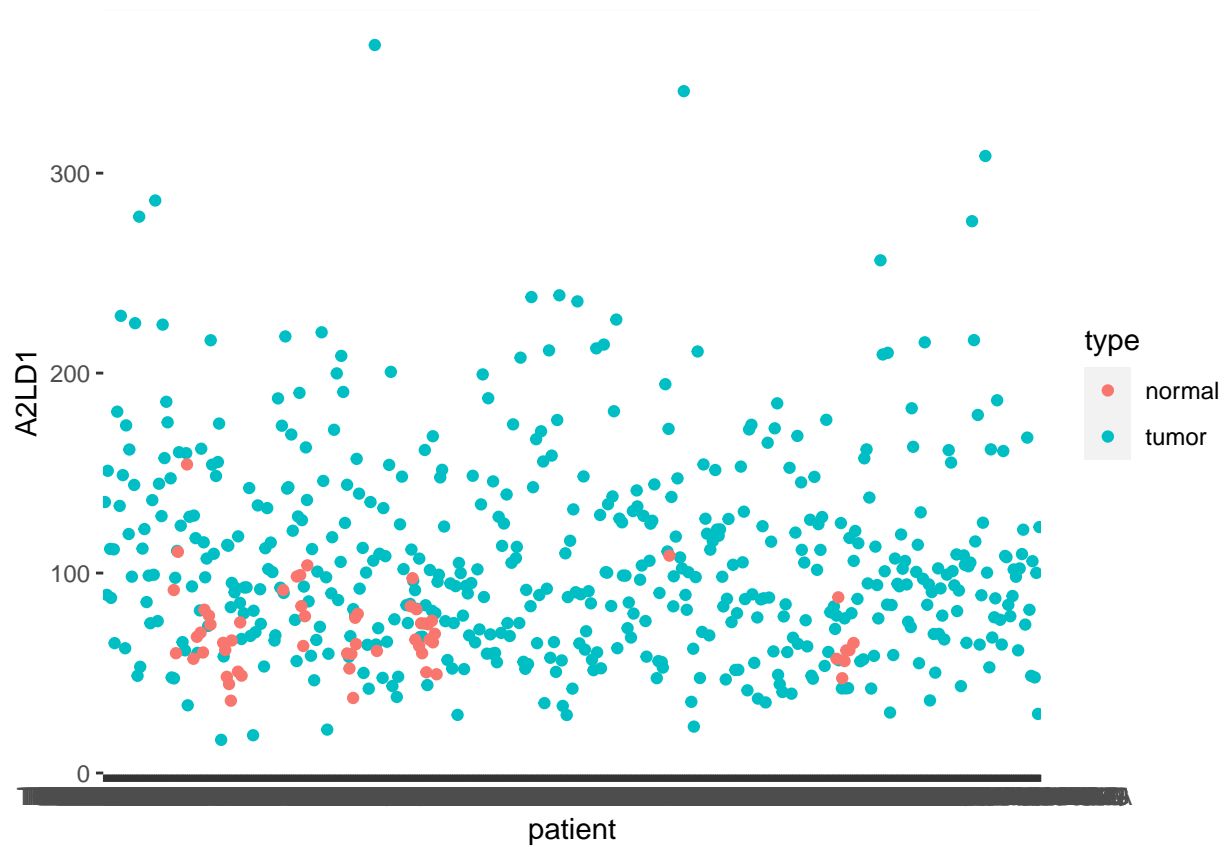
```



Question 2

Repeat this visualisation using this time the `geom_jitter`. Which representation is more appropriate? Why?

```
ggplot(expression, aes(x = patient , y = A2LD1, colour=type)) + geom_jitter()
```

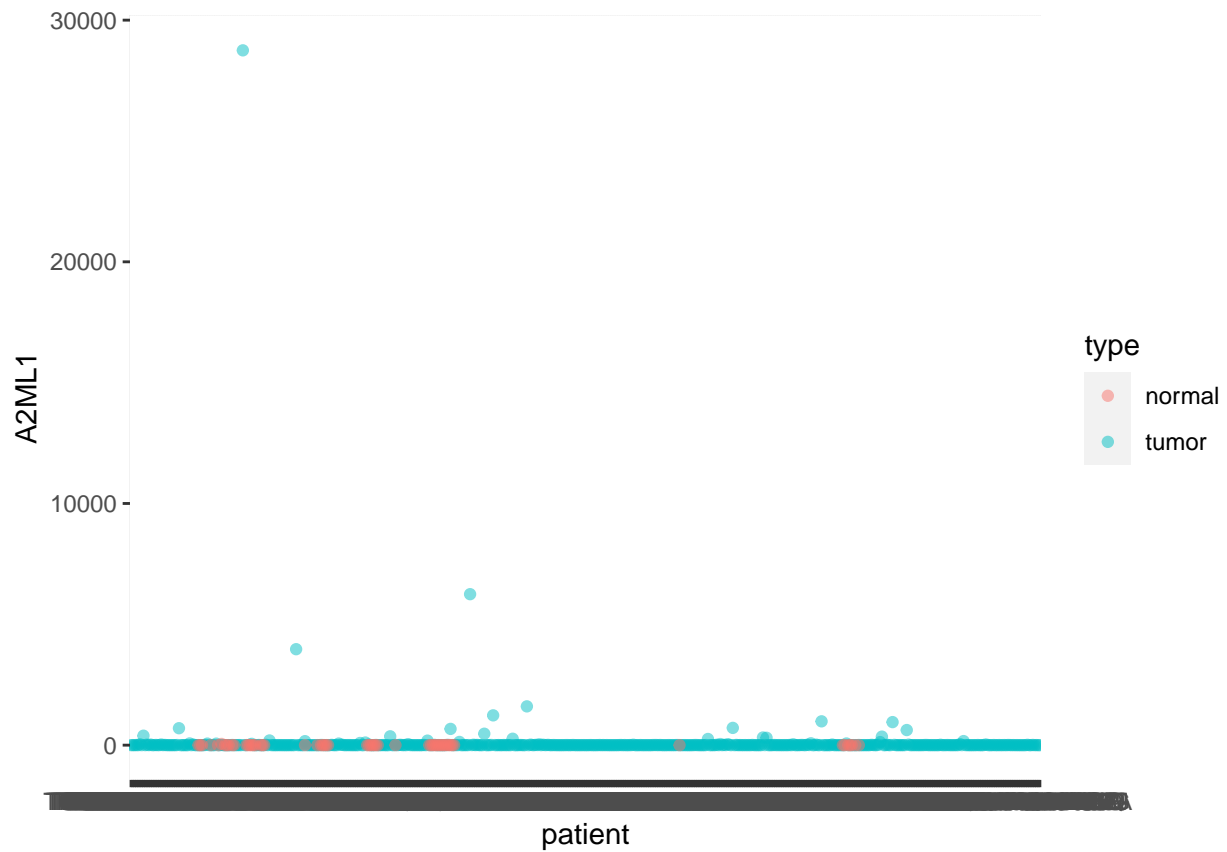


I've got the same results so idk dude

Question 3

Colour the samples according to their expression level in A2ML1.

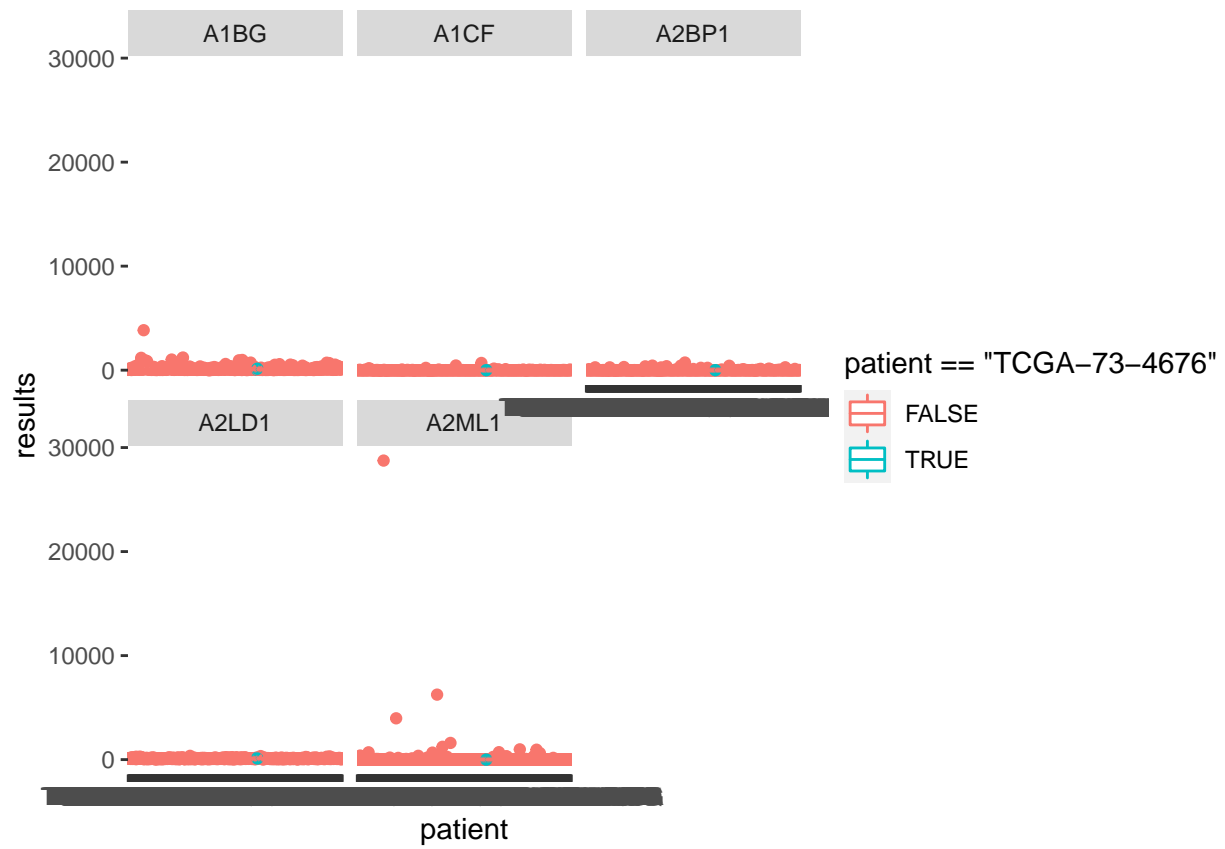
```
ggplot(expression, aes(x= patient, A2ML1, colour = type)) + geom_point(alpha=0.5)
```



Question 4

Highlight the points corresponding to patient "TCGA-73-4676."

```
lg <- pivot_longer(expression, cols = c(A1BG, A1CF, A2BP1, A2LD1, A2ML1), names_to = "genes", values_to = "results")
ggplot(lg, aes(x = patient, y = results, colour = patient == "TCGA-73-4676")) + geom_point() + geom_boxplot(aes(x = patient))
```



Question 5