Report Summary

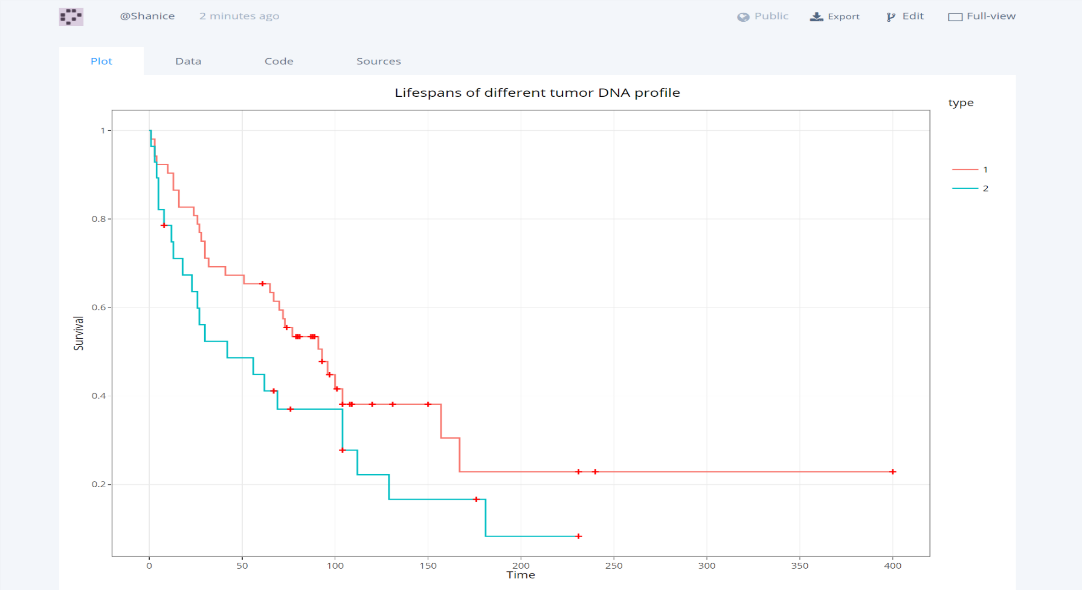
**Homework 1**

* Load the R code in your RStudio, and plot something nice (and relevant to survival analysis) and publish it on Plotly.
* Email URL to NU7250@gmail.com,along with R code and screenshot of graphs. Mention in the report about the dataset you printed and why.



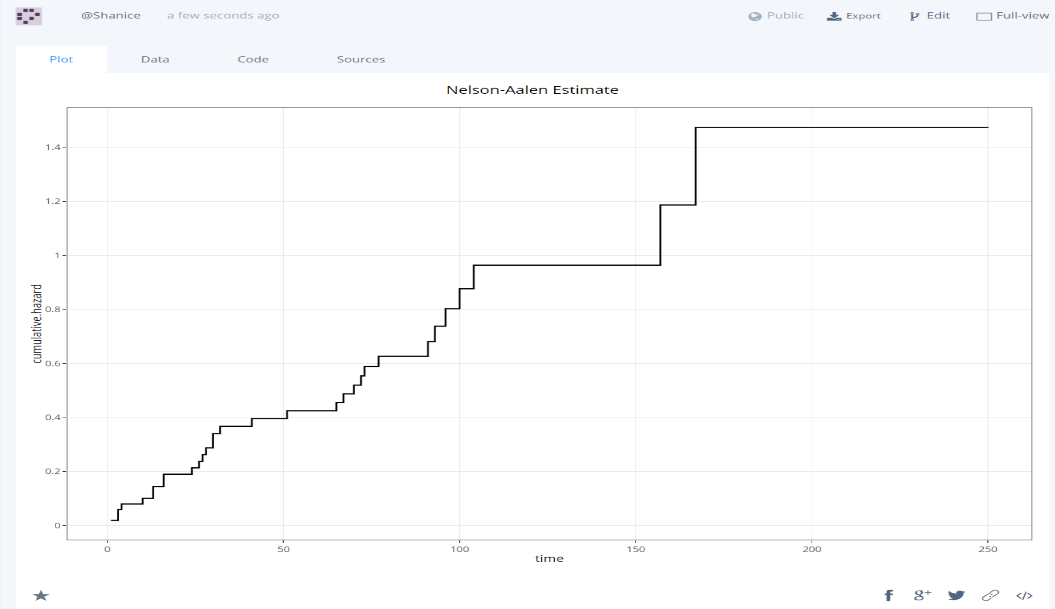
Kaplan-Meier Estimate

The chart is about the Time and Survival. And it shows that when time goes, the survival rate goes down and then tends to be stable.



The Lifespans of different tumor DNA profile

This chart shows different groups have different survival rate, and have a big difference between them, ad finally they all tends to be stable.



Nelson-Aalen Estimate

The Nelson–Aalen estimator is a non-parametric estimator of the cumulative hazard rate function in case of censored data or incomplete data.

**My data**

I used the hodg dataset, which is about Hodgkins disease death and time.



All chart in Plotly and the website:

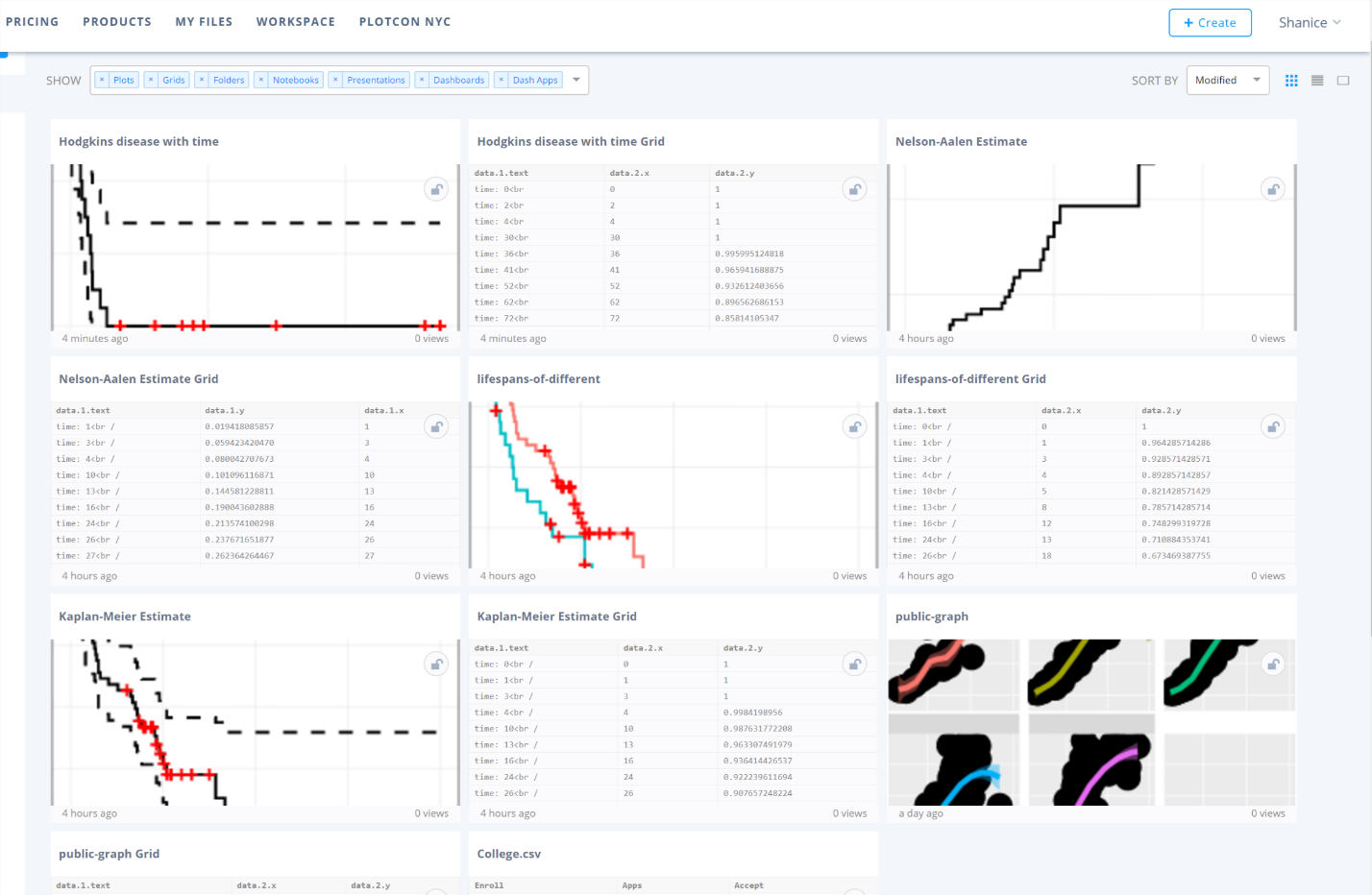
Chart 1 for Kaplan-Meier Estimate: <https://plot.ly/~Shanice/9>

Chart 2 for lifespans-of-different: <https://plot.ly/~Shanice/5>

Chart 3 for Nelson-Aalen Estimate: <https://plot.ly/~Shanice/7>

Chart 4 for Hodgkins disease Estimate: <https://plot.ly/~Shanice/11>

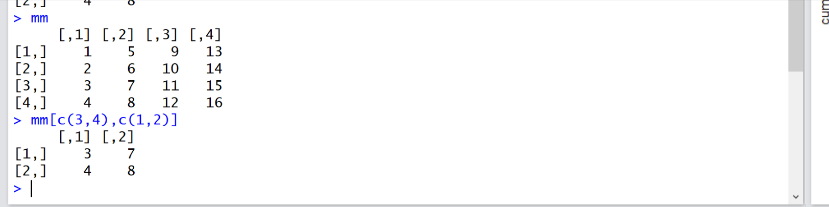
**Homework-2**

* Complete the R exercises. Exercises are present in 0-intro.R, 1-data.R.
* 0-intro.R

## exercise: select 2x2 subsection from the "bottom left" of matrix mm

## [your code here]

**mm[c(3,4),c(1,2)]**



* 1-data.R

## EXERCISE

## obtain this data view from "df":

## X Grad.Rate

## 1 James Madison University 98

## 2 Incarnate Word College 95

## 3 Johns Hopkins University 90

## 4 John Carroll University 89

## 5 Kenyon College 88

## 6 King's College 87

## 7 La Salle University 84

## 8 Illinois Wesleyan University 83

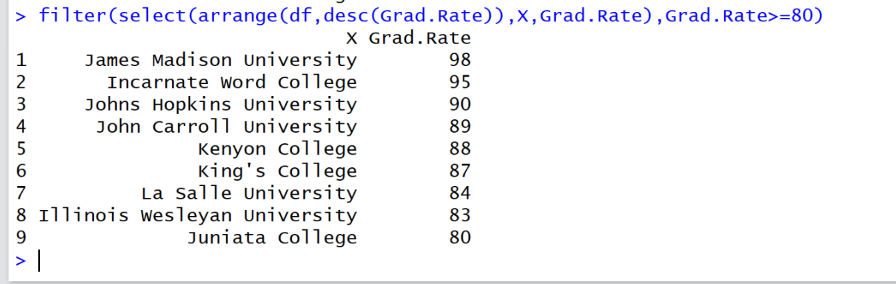
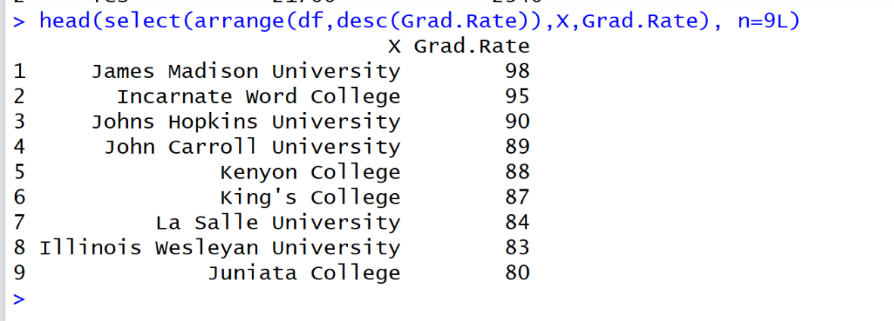
## 9 Juniata College 80

## [your code here]

**head(select(arrange(df,desc(Grad.Rate)),X,Grad.Rate), n=9L)**

##or

**filter(select(arrange(df,desc(Grad.Rate)),X,Grad.Rate),Grad.Rate>=80)**



## EXERCISE

## find max and min tuition ("Outstate") grouped by private/public

## school, in dataset 'df' and 'college'

## DF:

## Private max min

## 1 No 9766 3946

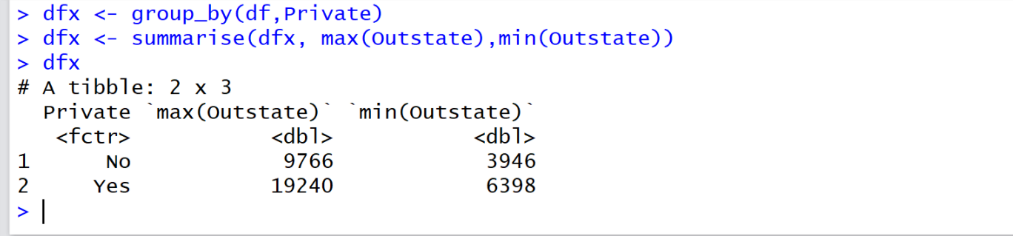
## 2 Yes 19240 6398

## [your code here]

**dfx <- group\_by(df,Private)**

**dfx <- summarise(dfx, max(Outstate),min(Outstate))**

**dfx**



## college:

## Private max min

## 1 No 15732 2580

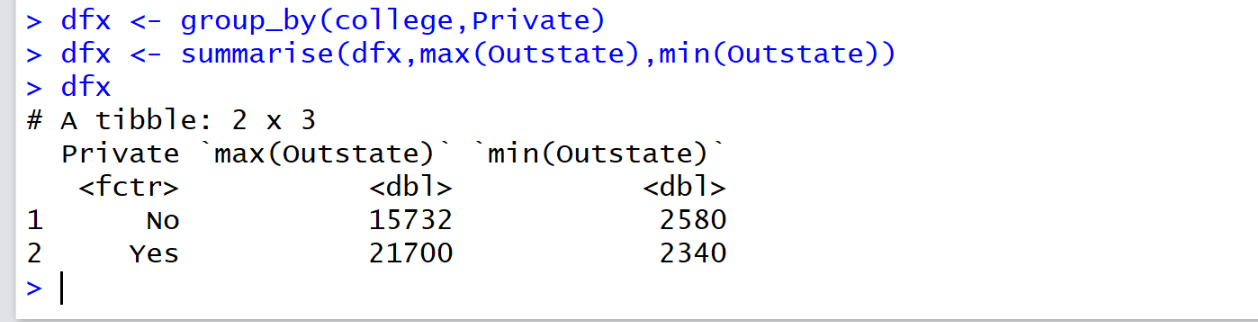
## 2 Yes 21700 2340

## [your code here]

**dfx <- group\_by(college,Private)**

**dfx <- summarise(dfx,max(Outstate),min(Outstate))**

**dfx**

****