

در **rstudio** سه عدد داده مربوطه را لود کرده و ستون اول را حذف میکنیم زیرا صرفاً کاربرد **annotation** دارد و در اینجا استفاده ای از آن نداریم.

توابع مختلفی را برای حل این سوال پیاده سازی میکنیم که هر کدام به شیوه ای به ما کمک میکنند تا مساله حل شود. تابع **get\_tissue\_from\_SRR** به کمک داده **leaf\_root\_annotation.csv** به ما کمک میکند تا با دانستن رکورد اسم **sample** اسم بافت را استخراج کنیم.

تابع **calc\_label\_df** کار اساسی را برای قسمت **a** سوال سوم انجام میدهد. با دریافت داده ورودی **raw** یا **normalized** و تعداد کلاسترهای از پیش تعیین شده، یک **knn** انجام میدهد و داده های **classification** این الگوریتم را همراه با داده های **tissue** و **project** از داده **annotation** را به خروجی میبرد و قسمت **a** سوال را حل میکند.

تابع **pca\_study** کار اساسی را برای قسمت **b** سوال سوم انجام میدهد. با دریافت داده اصلی و یک رشته **tag**، الگوریتم **principle component analysis** فراخوانی میشود و نتایج به وسیله **ggplot** ترسیم میگردد

خروجی قسمت اول یک دیتافرم هست که دسته بندی براساس **knn**، و داده های **annotation** را خلاصه بندی میکند. نحوه **visualization** و **correlation** هر کدام از لیبل ها در صورت سوال درخواست نشد. فلذا پیاده سازی نیز نگردید.

نتایج قسمت **a** در فایل **code\_log\_output.txt** و نتایج قسمت **b** در فایل **graphical\_results\_q3\_b.pdf** قابل مشاهده است