importDataAndgetInShape

This script loads all the R libraries and imports all data from SVN. It also converts the data to the correct formats. This script is sourced to the other scripts in the project, and thus, it always runs.

Import all files

Steps to take before importing to RStudio:

- Get data from qlikview (grades, enrolStatus) and from Cristina Draghici the kvote and GymData.
- With clickview data import into GoogleSheets go to bottom: REMOVE:::: Selection Status: Uddannelse: Medialogiand
- Export/download as CSV and move to SVN

Note: Grades B-passed, I-failed, EB - not graded (not allowed to sit exam), U - no show but their specific meanings regarding when you can take the re-exam.

Importing data from SVN

```
SVNData<-if(grepl("BiancaClavio", getwd()))
  {'C:/Users/BiancaClavio/Documents/SVN/01Projects/dropOut/data'}else
  {"~/01Projects/dropOut/data/"}

setwd(SVNData)
```

Setting the working directory to the SVN folder.

```
dfMed1Q999<-read.csv("Med1Q999.csv",
  header = TRUE, fill=TRUE, sep = ",",fileEncoding = "UTF-8")

dfMed1Interviews<-read.csv("Drop-out interviews - questionnaire.csv",
  header = TRUE, fill=TRUE, sep = ",",fileEncoding = "UTF-8", check.names=FALSE)

dfMed1Q999<-dfMed1Q999[,1:7]
```

Importing the interviews files with Danish encoding (UTF-8). When you read the data in then the names will not be changed (check.names=FALSE). Lastly, we only keep the first seven columns of the first variable ([,1:7]).

```
dfECTSstruct<-read.csv("course_SPV.csv",
  header = TRUE, fill=TRUE, sep = ",",fileEncoding = "UTF-8")
```

Importing all courses distributed on different study plan versions.

```
dfEnrolStatusMsc<-read.csv("MedEnrolMScThroughJul2017.csv",
  header = TRUE, fill=TRUE, sep = ",",fileEncoding = "UTF-8", check.names=FALSE)
dfEnrolStatusMsc$fradatosn<-as.Date(
  as.character(dfEnrolStatusMsc$fradatosn), "%d.%m.%Y")
dfEnrolStatusMsc$stype<-as.factor("kandidat")

dfEnrolStatusBsc<-read.csv("MedEnrolBScThroughJul2017.csv",
  header = TRUE, fill=TRUE, sep = ",",fileEncoding = "UTF-8", check.names=FALSE)
dfEnrolStatusBsc$fradatosn<-as.Date(
  as.character(dfEnrolStatusBsc$fradatosn), "%d.%m.%Y")
```

```
dfEnrolStatusBsc$stype<-as.factor("bachelor")
dfEnrolStatusBsc$Studieordningskode<-as.factor(NA)
```

Importing the enrolment status of all bachelor and master students on Medialogy from 2011 to 2017. The enrollment date at the education (fradatosn) for each education is converted to dates types. *stype* stores the type of education, and it coerces its argument to a factor (i.e. "kandidat" and "bachelor").

```
dfEnrolStatus<-rbind(dfEnrolStatusMsc,dfEnrolStatusBsc)
dfEnrolStatus<-dfEnrolStatus[!(dfEnrolStatus$udmeldsn
  =='Afvist på grund af manglende kvalifikationer'),]
dfEnrolStatus$enrolID<-seq(1:nrow(dfEnrolStatus))
dfEnrolStatus$navn<-NULL
```

Afterwards, the enrolment status of all bachelor and master students are combined by columns or rows (rbind). Students dropping out due to a lack of qualifications are removed. A enrolment id is generate in a regular sequence. Lastly, all names of the students are removed from the dateset.

```
dfEnrolStatus$fradatosn<-as.Date(as.character(dfEnrolStatus$fradatosn)
  , "%d.%m.%Y")
dfEnrolStatus$slutdatosn<-as.Date(as.character(dfEnrolStatus$slutdatosn)
  , "%d.%m.%Y")
dfEnrolStatus$yearOfEnrolment <- dfEnrolStatus$startaar
dfEnrolStatus$startaar<-as.numeric(as.character(dfEnrolStatus$startaar))
```

The combined enrolment status is converted to dates types for the enrolment date and the last day at the education. The variable of start years is stored in another variable and is