LOAD ERROR RPY2 (Windows10)

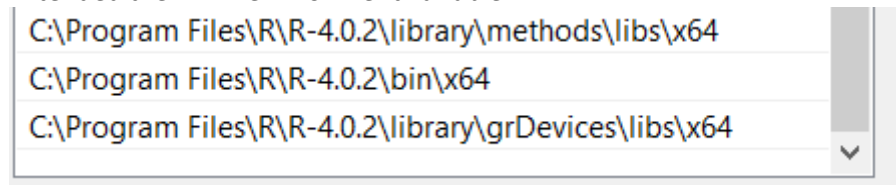18, 19 January 2021
CS109-2015 Lab11

Needed 3 files: model_matrix.csv and target.csv are nowhere to be found. File food_inspections.csv can be downloaded (e.g. from Kaggle) but I prefer to have all 3 based on the Chicago set used in the CS109-2015 lab.

So decided to generate the CSV file by conversion from the original Rds files using R-magic in Python Jupyter Notebook `lab11-2-Data-Processing.ipynb`
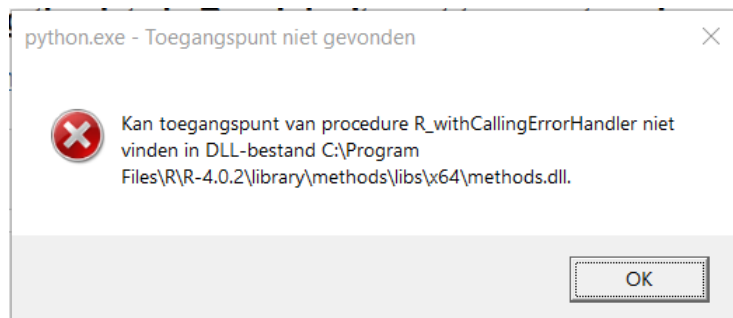
RPY2 installed with `pip` => errors on `%load_ext rpy2.ipython`
RPY2 installed via `conda install -c conda-forge rpy2` => errors op `%load_ext rpy2.ipython`

Extended the PATH environment variable:

```
C:\Program Files\R\R-4.0.2\library\methods\libs\x64
C:\Program Files\R\R-4.0.2\bin\x64
C:\Program Files\R\R-4.0.2\library\grDevices\libs\x64
```

resulting in errors below:

python.exe - Toegangspunt niet gevonden

Kan toegangspunt van procedure R_withCallingErrorHandler niet vinden in DLL-bestand C:\Program Files\R\R-4.0.2\library\methods\libs\x64\methods.dll.

OK

python.exe - Toegangspunt niet gevonden

Kan toegangspunt van procedure Rf_translateCharFP niet vinden in DLL-bestand C:\Program Files\R\R-4.0.2\library\grDevices\libs\x64\grDevices.dll.

OK

The 'Anaconda prompt' terminal displays:
During startup - Warning messages:
1: package 'datasets' was built under R version 4.0.2
2: package 'utils' was built under R version 4.0.2
3: package 'grDevices' was built under R version 4.0.2
4: package 'grDevices' in options("defaultPackages") was not found
5: package 'graphics' was built under R version 4.0.2

6: package 'graphics' in options("defaultPackages") was not found
7: package 'stats' was built under R version 4.0.2
8: package 'stats' in options("defaultPackages") was not found
9: package 'methods' was built under R version 4.0.2
10: package 'methods' in options("defaultPackages") was not found

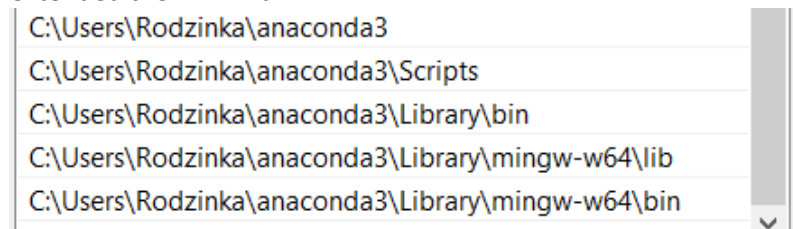Based on https://anderfernandez.com/en/blog/how-to-program-with-python-and-r-in-the-same-jupyter-notebook/
these environment vars are added:

R_HOME	C:\Program Files\R\R-4.0.2
R_LIBS_USER	C:\Users\Rodzinka\Documents\R\win-library\4.0
R_USER	C:\Users\Rodzinka\anaconda3\Lib\site-packages\rpy2

Restarting Jupyter notebook (kernel restart and clean) gives same errors.

According to https://anaconda.zendesk.com/hc/en-us/articles/360023857134-Setting-up-rpy2-on-Windows
extended the PATH var:

```
C:\Users\Rodzinka\anaconda3
C:\Users\Rodzinka\anaconda3\Scripts
C:\Users\Rodzinka\anaconda3\Library\bin
C:\Users\Rodzinka\anaconda3\Library\mingw-w64\lib
C:\Users\Rodzinka\anaconda3\Library\mingw-w64\bin
```

After another Jupyter notebook restart the same errors as above.

After restarting computer:

```
%load_ext rpy2.ipython

C:\Users\Rodzinka\anaconda3\lib\site-packages\rpy2\robjects\packages.py:366: UserWarning: The symbol 'quartz' is not
in this R namespace/package.
  warnings.warn(
```

which indicates a successful load (with a warning that I looked up but ignored).

Now the next notebook cell executes under R-magic

```
%%R
# change to your local clone
data_dir = '~/food-inspections-evaluation/'
out_dir = '~'

library("data.table", "ggplot2")

setwd(data_dir)
```
```
R[write to console]: Error in library("data.table", "ggplot2") :
  there is no package called 'data.table'
```

and displays an error of some missing package.

OK, installed this package from RStudio as follows: `install.packages("data.table", dependencies=TRUE)`

So finally could finish conversion to the 3 CSV files needed in Lab11 part 1.