



PyData
Bratislava



Power of Python, R & Tableau in Data Science

(PyData Bratislava Meetup #1, Nervosa; #PyDataBA)

13. 2. 2017

Radovan Kavicky, GapData Institute (GDI)



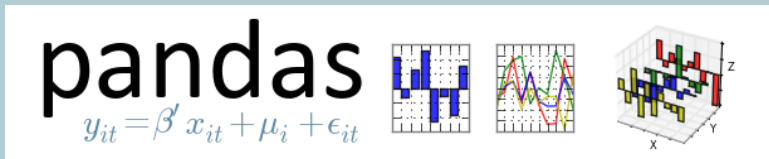
GapData Institute (GDI)



- Economic Research & Public Policy & Data Science think-tank (data-tank)
- Data. Think. Change.
- GapData Institute (GDI) is a non-profit nonpartisan research institution harnessing power of data & wisdom of economics for public good.
- Transparent account (from day #1; SK7383300000002200933920 <https://www.fio.sk/ib2/transparent?a=2200933920>)
- Partnership (openness, transparency)



Python (best tools/Data Science)



- ▣ pandas (data manipulation, analysis)
 - pandas datareader (data import)
 - pandas-ply (functional data manipulation)
 - datacleaner (automate clean your data)
- ▣ matplotlib (dataviz)
- ▣ SciPy (scientific Python)
- ▣ NumPy (numerical Python)
 - Numba (app high-perf)
- ▣ Bokeh (dataviz)
- ▣ scikit-learn (ML)





R (best tools/Data Science)



- ▣ R-Studio (IDE for R)
 - Shiny (R apps/interactive tools)
- ▣ ggplot2 (dataviz, plotting tool)
- ▣ ggvis (dataviz, declarative describe)
- ▣ dplyr (data manipulation, dataframe)
- ▣ data.table (data aggregator, join+modify)
- ▣ knitr (report generation)





Tableau (best tools/Data Science)



- ▣ Tableau Desktop (+Public Edition)
- ▣ Tableau Server (Python +R models)
- ▣ Power Tools for Tableau
 - Power Tools for Tableau Desktop
 - Power Tools for Tableau Server
 - Power Tools for Tableau Deployment
- ▣ Workbook SDK for Tableau (soft dev)
- ▣ DataWrangler (wrangle data took)
- ▣ Alteryx Visual Analytics Kit for Tableau



Python, R & Tableau (how to connect them)



- ▣ IPython/Project Jupyter (Julia, Python + R)
- Jupyter notebook (.ipynb rendering via web browser) +nbviewer

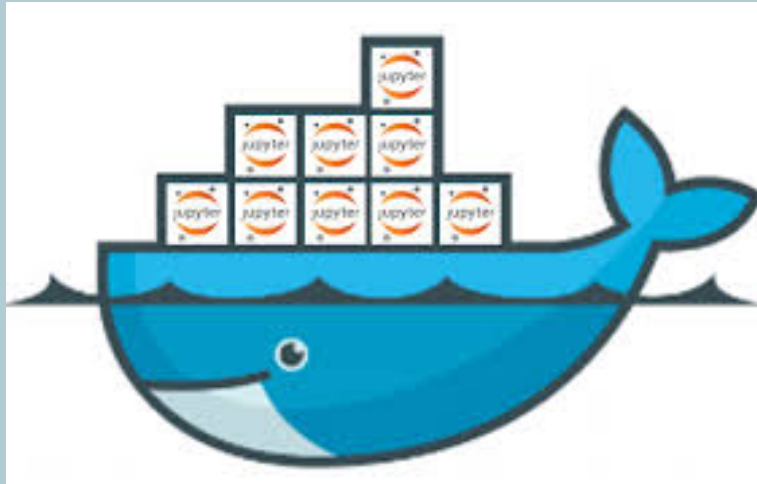


- ▣ rpy2 (start R in Python)
- ▣ TabPy (connect to Tableau Server from Python) +dplython (dplyr @Py)
- ▣ D3 (DataViz, JavaScript)
- ▣ Tableau +R.Data (connect .Rproj +.RData)
- ▣ PythonInR (start Python in R)
- ▣ Rserve (TCP/IP binary R server)





Data Science (other tools)



- ❑ JupyterHub (Jupyter for teams)
- ❑ Docker (virtualization, app deploy)
- ❑ Kafka (distributed stream platform)
- ❑ Hadoop (Apache + distributed data storage + MapReduce, Java)
- ❑ MongoDB (cross-platform document-oriented database)
- ❑ Apache Spark (cluster computing)
- ❑ +many open-source tools





Other activities (GapData Institute)

- ▣ PyData Bratislava, R <- Slovakia, skczTUG
- ▣ Open Data (only the 1st necessary step)
- ▣ Python, R, Tableau community activities
- ▣ Economic Research (studies, papers, publications)
- ▣ Quantitative & Qualitative Analyses (Data/center of it all)
- ▣ Data Visualization (Interactive DataViz Tools)
- ▣ Public Policy (Public Budget, Public Finance)
- ▣ Economic Reforms (any area)
- ▣ PyCon SK 2017 (10th-12th of March 2017, FIIT UK)



Thank you for your attention

Contact:

Radovan Kavicky



radovan.kavicky@gapdata.org



+420 776 467 159 (CZ)

+421 949 716 214 (SK)



<http://www.linkedin.com/in/radovankavicky>



<https://gapdata.slack.com/messages/py-data/>



<https://github.com/radovankavicky>

<https://github.com/GapData/PyDataBratislava>



@radovankavicky, @PyDataBA, @GapDataInst



#PyDataBA

In case you have any question, feel free to ask.