Prof. Dr. Thomas Schultz
Mohammad Khatami (khatami@cs.uni-bonn.de)
Ikram Jumakulyyev (ijumakulyyev@cs.uni-bonn.de)

University of Bonn
B-IT/Computer Science II
April 21, 2020

Summer term 2020

Visual Data Analysis

Assignment Sheet 1

Solution has to be sent until April 27, 2020, 8:00am
to khatami@cs.uni-bonn.de **and** ijumakulyyev@cs.uni-bonn.de.

Please bundle the results (as PDF) and scripts (*.py/*.ipynb files) in a single ZIP file.

Include your names explicitly in the PDF and each script.

Please send a pdf that contains the names of your team members, your answer to the three questions below, and a screenshot of the visualization you created to answer them, until **April 27, 2020, 8:00am** to khatami@cs.uni-bonn.de **and** ijumakulyyev@cs.uni-bonn.de. Please put [**vda2020-ex01**] into the subject of your email.

Help with Installing the Required Software
If you encounter problems with your installation, please subscribe and write to our mailing list, vl-scivis@lists.iai.uni-bonn.de. There will *not* be a regular exercise class on April 29, 2020. However, a TA will be there to provide help with any installation issues that might remain. Since the hand-in date for this sheet is before that, it only contains bonus points, none that you need to be admitted to the exam.

Exercise 1 (Install and try Python, *10 Bonus Points*)

The programming tasks in this lecture will be in Python. Doing them requires some open source software, which is available for Linux, Windows, and Macs.

- If you are new to Python, one option to install the required software is via the Python 3.7 version of the Anaconda distribution, which you can download for free from `https://www.anaconda.com/distribution/`.
- If you already have a recent version of Python, you should be able to continue using your preferred environment, which should include the Python interpreter itself (Python 3 strongly recommended), and recent versions of packages such as NumPy, SciPy, matplotlib, pandas, and seaborn.

Jupyter notebook is quite useful for running quick experiments with Python. For writing scripts, we recommend using an integrated development environment that supports Python. One possible choice is Spyder. If you decide to install Anaconda, all this software will be included by default.

Seaborn provides a simple way to plot data in Python. Please follow the first chapter ("Visualizing statistical relationships") of the official seaborn tutorial at `https://seaborn.pydata.org/tutorial.html`. Start playing around with this software. Based on the "tips" dataset introduced in the tutorial, create a visualization that allows you to answer the following three questions:

a) Are there days of the week on which the dataset covers only one meal, i.e., either lunch or dinner?
b) Was the highest tip given by a man or a woman?

c) Was the highest tip given for the highest total bill?

Hint: In case you require a more detailed introduction to Python itself, please refer to the Python tutorial at https://docs.python.org/3/tutorial/.

Good Luck!