



DIS08 – Data Modeling

03 – The Shell – Tutorial Session

Fabian Haak,
Technische Hochschule
Köln, Cologne, Germany

Version: 2021-10-22

Technology
Arts Sciences
TH Köln

Open Questions



Quiz Time!



Quiz:

<https://quizizz.com/join?gc=612034>

Hands-on Shell



Hands-on Shell

- The dataset we work with is available in Moodle (shell-lesson.zip). Download and unzip in a directory of your choice. (you can use your shell to unzip the file! → „unzip shell-lesson.zip“ in the file's directory)
- **please don't add it to git.**
- The dataset mimics some real-life task: You have been handed some files, and your task as a data scientist is to inspect them before you can do some analysis.

Hands-on Shell

- Some example tasks:
 - List the content of the directory. List all files in long format, with human-readable size. Order files by timestamp. Order files by size.
 - Have a look at the files' contents. View the complete content, or just peak at the first few lines. Control how many lines you will see. Have a look at the last few lines to see how a long file ends.
 - Count the number of lines, characters or words in a file.
 - Count the number of lines, characters or words for all files, sort them by one of the counts.
 - When working with CSV files (file extension .csv or .tsv): Have a look at the header, i.e. column names. Store the column names to another file for reference.
 - Create a new directory for backups. Move files to that directory. Remove files. Remove the whole directory.

Hands-on Shell

Some example commands:

```
ls
```

```
ls -a
```

```
ls -lh
```

```
ls -alht
```

```
cat diary.html
```

```
less diary.html
```

```
head diary.html
```

```
head -n 5 diary.html
```

```
tail diary.html
```

```
head -n 1 2014-01_JA.tsv
```

```
head -n 1 2014-01_JA.tsv > header.csv
```

```
head -n 1 2014-02-02_JA-britain.tsv >> header.csv
```


Hands-on Shell

Some more example commands:

```
wc diary.html
```

```
wc *.tsv
```

```
wc -l *.tsv
```

```
wc -w *.tsv | sort
```

```
mkdir backup
```

```
mv header.csv backup/
```

```
rm backup/header.csv
```

```
rm -r backup/
```

```
rmdir backup/
```

Shell: Getting help and helping yourself

- Usually, each command has an entry in the manual. The manual explains the command's purpose, usage and options.
- In fact, the manual itself is a program, so there's a command `man` to access it.
 - Unfortunately, this doesn't work on git Bash for Windows.
 - There's also an online version, see References.
- Besides the manual, there's often a `--help` option available for each command.
- If you're stuck in some output (e.g. from `less`) or some process that won't stop, you can either use the **q** key, or **Ctrl+C** to exit.

Time for your Assignment and Questions



References

- Linux Man Pages Online:
<http://man.he.net>
- Add color to the terminal in MacOS:
<http://osxdaily.com/2012/02/21/add-color-to-the-terminal-in-mac-os-x/>