

# Final assignment for the course Data visualization with R

Master Life Science Informatics

## Methods:

- Methods from descriptive statistics: Descriptive Statistics
- Statistics for scientific work: Literate Statistical Analysis
- Advanced Data visualization techniques: ggplot2

## Introduction:

As an expert on Life Science Informatics, you are required to analyse the US weekly Nationally Notifiable Disease Surveillance Data from 1888 to 2013. The data can be used to estimate seasonal and long-term transmissions trends, generate models for predictability of infectious disease outbreaks and conduct scientific research work.

Based on the data you are expected to prepare a presentation (power point slides) conducting an Exploratory Data Analysis, for which you have 5 minutes and one R notebook with scientifically oriented analysis (2000-3000 words, including scientific literature). The level 2 data contains 50 diseases 50 US states and 1284 US cities, so you are free to choose the topic you want to work with.

## Data to be analysed:

Project Tycho (Source: <https://www.tycho.pitt.edu>)

## Resources:

VisGuides host a community of people working with the same data:  
<https://visguides.org/search?q=tycho>

## Criteria:

- Good research question and result
- Good contextualisation of the statistical analysis in the selected region
- Quality and diversity of the graphics used
- Quality of the presentation: engaging capacity and clarity of the message.
- Using statistical key figures correctly
- Use and explanation of (advanced) statistical and visualization techniques not covered in the course.

## Tools:

R and Rstudio, any other additional dataset you may consider relevant.

# Formal requirements for the final assignment project

## **Title page**

Logo, subject, subject, mat no., e-mail, name

## **Overview**

Short summary, keywords

List of contents, figures, tables, listings, abbreviations ...

### **1. Introduction**

Short description of the topic and transition to the Problem/State of the Art.

### **2. Problem definition**

State-of-the-Art

Introduction to the problem

### **3. Objectives**

Brief formulation of the objectives of this work.

### **4. Methods**

Description of the method and different Steps, what has been done? Why it is done in that way?

### **5. Results**

Presentation and interpretation of the results.

### **6. Discussion**

Pro- and contra discussion on results and the work.

### **7. Conclusions**

Concluding words and a brief outlook on what else could be done or what the results could be used for.

## **Literature**

Citations, online sources, data, images, ...

## **Appendix**

Maps, source code, files...

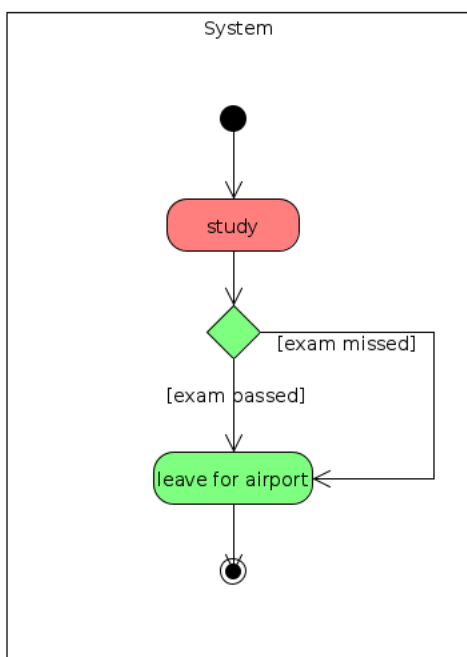
## Tables

**Tab. 1:** Frei erfundene Daten. Daten aus [1].

Col 1	Col 2	Col 3	Col 4
50,004	34,986	45,367	56,38
45,095	56,378	93,387	7,378
78,092	76,456	8,738	398,000

- Cite external data!
- Each table has a heading!
- Own data is not quoted!
- Foreign and own data are not screened!
- Tables in computer science ALWAYS have this format!
- Tables are referenced and mentioned from the text (min. 4 to 5 sentences per table) (Tab. 1)

## Figures



**Fig. 1:** Any picture. Data from [2].

- Images always have the title under the figure
- If graphics are rebuilt, the original must be quoted
- Own graphics are always (very) positive!
- Graphics should not be blurred (Everything well readable: 320 dpi)
- Graphics are referenced and mentioned from the text (min. 4 to 5 sentences per Figure) e.g. (Fig. 1)

## Listings

**Lst. 1:** Python Hallo-Welt Programm. Quellcode aus [1].

---

```
0 #!/usr/bin/env python3
1
2 def main():
3     print("Hallo Welt")
4
5 if __name__ == '__main__':
6     main()
```

---

- Quote foreign code snippets and programs!
- Every listing has a headline!
- Own programs are not quoted! These will be added to the appendix and it will be this one! Also too long programs are not used for listing!
- Foreign and own source codes are NEVER screen-shoted!
- Listings are referenced and mentioned from the text (min. 1 to 3 sentences per Listing) (Lst. 1)!

## Citations:

**Web links and URLs:** All web links and URLs, including links to the authors' own websites, should be given a reference number and included in the reference list rather than within the text of the manuscript. They should be provided in full, including both the title of the site and the URL, as well as the date the site was accessed, in the following format: The Mouse Tumor Biology Database. <http://tumor.informatics.jax.org/mtbwi/index.do>. Accessed 20 May 2013.

### Examples reference style:

#### *Article within a journal*

Smith J, Jones M Jr, Houghton L (1999) Future of health insurance. N Engl J Med 965:325-329.

#### *Article by DOI (with page numbers)*

Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. J Mol Med 78:74-80. doi:10.1007/s001090000086.

#### *Article by DOI (before issue publication and with page numbers)*

Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. J Mol Med. doi:10.1007/s001090000086.

#### *Article in electronic journal by DOI (no paginated version)*

Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. Dig J Mol Med. doi:10.1007/s801090000086.

More information on citations in: <https://energyinformatics.springeropen.com/submission-guidelines/preparing-your-manuscript/research>