4th ECPR Winter School in Methods and Techniques, 13-20 February 2015
University of Bamberg, Germany
Course Description Form¹ [1-week main course, 15 hours]

Course title
WD101 Quantitative Text Analysis

Instructor details
First name, last name: Heike Klüver
Department/Unit: University of Bamberg
Institution: Professorship of Empirical Political Science
Full postal address for ECPR correspondence: Feldkirchenstrasse 21, 96045 Bamberg, Germany
Phone: +49 (0) 951-863 2493
Fax: E-mail : heike.kluever@uni-bamberg.de

Short Bio
Heike Klüver is Professor of Empirical Political Science at the University of Bamberg. Before coming to Bamberg, she worked as Juniorm professor at the University of Konstanz and as a Postdoctoral Research Fellow at Nuffield College at the University of Oxford. Heike Klüver received her PhD from the University of Mannheim. Her research interests include European politics, interest groups, political parties, coalition governments and quantitative text analysis.

Prerequisite knowledge
Note from the Academic Convenors to prospective participants: by registering to this course, you certify that you possess the prerequisite knowledge that is requested to be able to follow this course. The instructor will not teach again these prerequisite items. If you doubt whether you possess that knowledge to a sufficient extent, we suggest you contact the instructor before you proceed to your registration.

The following skills are helpful though not necessarily required to be able to follow the course:
- Familiarity with the R statistical software package
- Basic knowledge of the STATA statistical software package
- Basic knowledge of statistical analysis
- Familiarity with a Text editor and with the handling of text files

Short course outline
This applied course will provide you with an overview of quantitative text analysis methods that allow you to systematically extract information from political texts. The course will start with more traditional approaches such as manual hand-coding, but quickly moves to recent advances in political methodology that treat words as data. The course will begin with important concepts in content analysis such as content validity and intercoder reliability. We will afterwards take a closer look at manual hand-coding approaches before turning to computer-assisted dictionary-based text analysis techniques. This will be followed by a discussion of Wordscores and Wordfish, two cutting-edge content analysis approaches that allow you to automatically extract policy positions from political

¹ Disclaimer: the information contained in this course description form may be subject to subsequent adaptations (e.g. taking into account new developments in the field, specific participant demands, group size etc.). Registered participants will be informed in due time in case of adaptations.
texts. Finally, we will cover automated document classification techniques which allow for automatically classifying texts into different thematic areas. The course will combine theoretical sessions with practical exercises to allow participants to immediately apply the presented techniques.

**Long course outline**

This applied course will provide you with an overview of quantitative text analysis methods that allow you to systematically extract information from political texts. The course will start with more traditional approaches such as manual hand-coding, but quickly moves to recent advances in political methodology that treat words as data. The course will begin with important concepts in content analysis such as content validity and intercoder reliability. We will then take a closer look at manual coding approaches as for instance employed in the famous Comparative Manifesto Project which rely on human coders to code the content of texts according to a predefined category scheme. Afterwards, we will move to automated text analysis techniques by first discussing computer-assisted dictionary-based text analysis. Dictionary-based content analysis employs computers to code the content of documents by relying on a humanly devised codebook which assigns individual words to specific thematic categories. Next, we will deal with fully computerized text analysis techniques. We will first deal with Wordscores and Wordfish, two cutting-edge techniques that allow you to automatically extract policy positions from political texts such as election manifestos or speeches. Finally, we will cover automated document classification approaches which allow for automatically classifying texts into different thematic areas. For instance, using such document classification techniques, researchers can automatically classify thousands of texts such as press releases or laws into different policy areas. This course is an applied course for beginners and intermediate users of content analysis that provides participants with an overview of the theoretical foundations of quantitative text analysis, but which is mainly practical and applied so that participants learn how to use these methods in their own research. The course will therefore combine theoretical sessions with practical exercises to allow participants to immediately apply the presented techniques.

**Day-to-day schedule (Monday 16 February to Friday 20 February)**

<table>
<thead>
<tr>
<th>Day</th>
<th>Topic(s)</th>
<th>Details [NB : incl. timing of lecture v/s lab or fieldwork etc. hours]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>Introduction / Hand-Coding</td>
<td>Two 90min lectures</td>
</tr>
<tr>
<td>Day 2</td>
<td>Dictionary Coding / Dictionary Coding Exercise</td>
<td>90 min lecture, 90 min lab</td>
</tr>
<tr>
<td>Day 3</td>
<td>Wordscores / Words as data exercise</td>
<td>90 min lab, 90 min lecture</td>
</tr>
<tr>
<td>Day 4</td>
<td>Wordfish / Ideal point measurement exercise</td>
<td>90 min lecture, 90 min lab</td>
</tr>
<tr>
<td>Day 5</td>
<td>Document classification / Classifying documents exercise</td>
<td>90 min lecture, 90 min lab</td>
</tr>
</tbody>
</table>

**Day-to-day reading list**

For the precise literature references, see reference list below.

<table>
<thead>
<tr>
<th>Readings (please list at least the compulsory reading for the scheduled day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
</tr>
<tr>
<td>Day 2</td>
</tr>
<tr>
<td>Day 3</td>
</tr>
<tr>
<td>Day 4</td>
</tr>
<tr>
<td>Day 5</td>
</tr>
</tbody>
</table>
**Software and hardware requirements**

**Software programme**
- R
- STATA
- Yoshikoder
- JFreq

**Hardware requirements**
- No specific requirements

**Literature**


**Lecture room requirement**

For the exercise sessions, I need a computer lab.

**Preferred time slots**

I prefer to teach in the morning

**Other recommended courses (before or after this course)**

**Before this course:**

<table>
<thead>
<tr>
<th>Course title</th>
<th>Summer School</th>
<th>Winter School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction to R</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Webscraping with R (I will send you a suggestion for this course soon)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**After this course:**

<table>
<thead>
<tr>
<th>Course title</th>
<th>Summer School</th>
<th>Winter School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction to Python</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Webscraping with R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>