Association between status of a language and its number of basic colour terms

Hizniye Isabella Boga

12 July, 2017

Structure

Basic Colour Terms

What are Basic Colour Terms?

- monolexemic (no "lemon-colored" or "blue-green")
- no subsets of other terms ("crimson" $\Rightarrow$ "red")
- not restricted to narrow class of objects ("blonde")
- psychologically salient (most people won't come up with "scarlet" when shown a color)

Framework of Berlin and Kay's Study

- comparison of color term vocabulary in 98 languages by showing colored chips to participants
- participants were asked to name the colors of their language, and to select the best fits on a chart of 320 colors
- they were then asked to classify the same 320 color chips isolatedly, resulting in a map of color categories

Results

- the number of basic colour terms is between 2 and 12
- if there are 11 (like in most major languages), they correspond exactly to the English ones
- 11 basic color terms in English: red, orange, yellow, green, blue, purple, pink, brown, grey, black and white
- 12 colors: Russian (2 blues), Hungarian (2 reds)
- clear restrictions if there are fewer than 11 colors:
  - (1): **BLACK**and **WHITE** always exist.
  - (2): If there is a third one, it is **RED**.
  - (3): The fourth color term is either **GREEN** or **YELLOW** .

Evolution of Colour Vocabularies

- Fixed Sequence of evolutionary stages
- languages all started with two color terms
- new color terms are added later
- basic color terms are not lost in language change
- 11 or 12 color terms is the maximum

Evolution of Basic Colour Terms

- A 7-Stage system
- **STAGE I: BLACK AND WHITE**
- **STAGE II: RED**
- **STAGE III: GREEN OR YELLOW**
- **STAGE IV: GREEN OR YELLOW**
- **STAGE V: BLUE**
- **STAGE VI: BROWN**
- **STAGE VII: PURPLE, PINK, ORANGE AND GREY**

- when colour lexicon expands beyond seven terms $\rightarrow$ rapid expansion to full roster of eleven basic colour categories

Results in the following 7 Universals

1. *All languages contain terms for white and black.*
2. *If a language contains three terms, then it contains a term of red.*
3. *If a language contains four terms then it contains a term for either green or yellow (but not both).*
4. *If a language contains five terms, then it contains terms for both green and yellow.*
5. *If a language contains six terms, then it contains a term for blue.*
6. *If a language contains seven term, then it contains a term for brown.*
7. *If a language contains eight or more terms, then it contains a term for purple, pink, orange and grey, or some combination of these.*

The Data - Analysis, Results and further Research questions

Databases

- Data from two databases
- Ethnologue and WALS
- Ethnologue provided Status of Language
- WALS has four chapters on Colour Terms
  - Number of Non-Derived Basic Colour Categories
  - Number of Basic Colour Categories
  - Green and Blue
  - Red and Yellow

Ethnologue

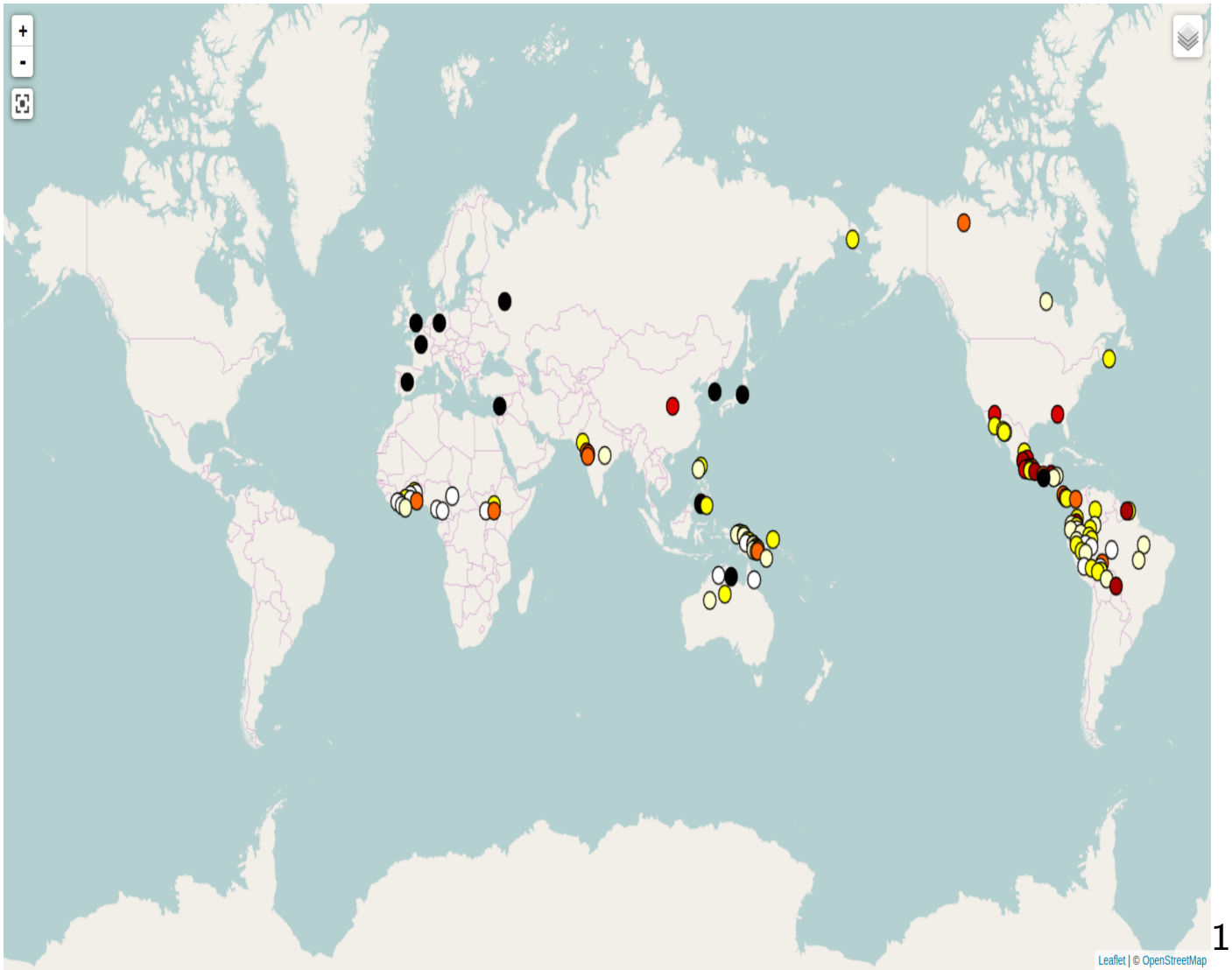
Ethnologue uses the following distinctive grading scale for the status of a language:

- **0** $\rightarrow$ INTERNATIONAL: The language is widely used between nations in trade, knowledge exchange, and international policy.
- **1**$\rightarrow$ NATIONAL: The language is used in education, work, mass media, and government at the national level.
- **2** $\rightarrow$ PROVINCIAL: The language is used in education, work, mass media, and government within major administrative subdivisions of a nation.
- **3** $\rightarrow$ WIDER COMMUNICATION: The language is used in work and mass media without official status to transcend language differences across a region.

- **4** $\rightarrow$ EDUCATIONAL: The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education.
- **5** $\rightarrow$ DEVELOPING: The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable.
- **6a** $\rightarrow$ VIGOROUS: The language is used for face-to-face communication by all generations and the situation is sustainable.
- **6b** $\rightarrow$ THREATENED: The language is used for face-to-face communication within all generations, but it is losing users.

- **7** → SHIFTING: The child-bearing generation can use the language among themselves, but it is not being transmitted to children.
- **8a** → MORIBUND: The only remaining active users of the language are members of the grandparent generation and older.
- **8b** → NEARLY EXTINCT: The only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language.
- **9** → DORMANT: The language serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency.
- **10** → EXTINCT: The language is no longer used and no one retains a sense of ethnic identity associated with the language.

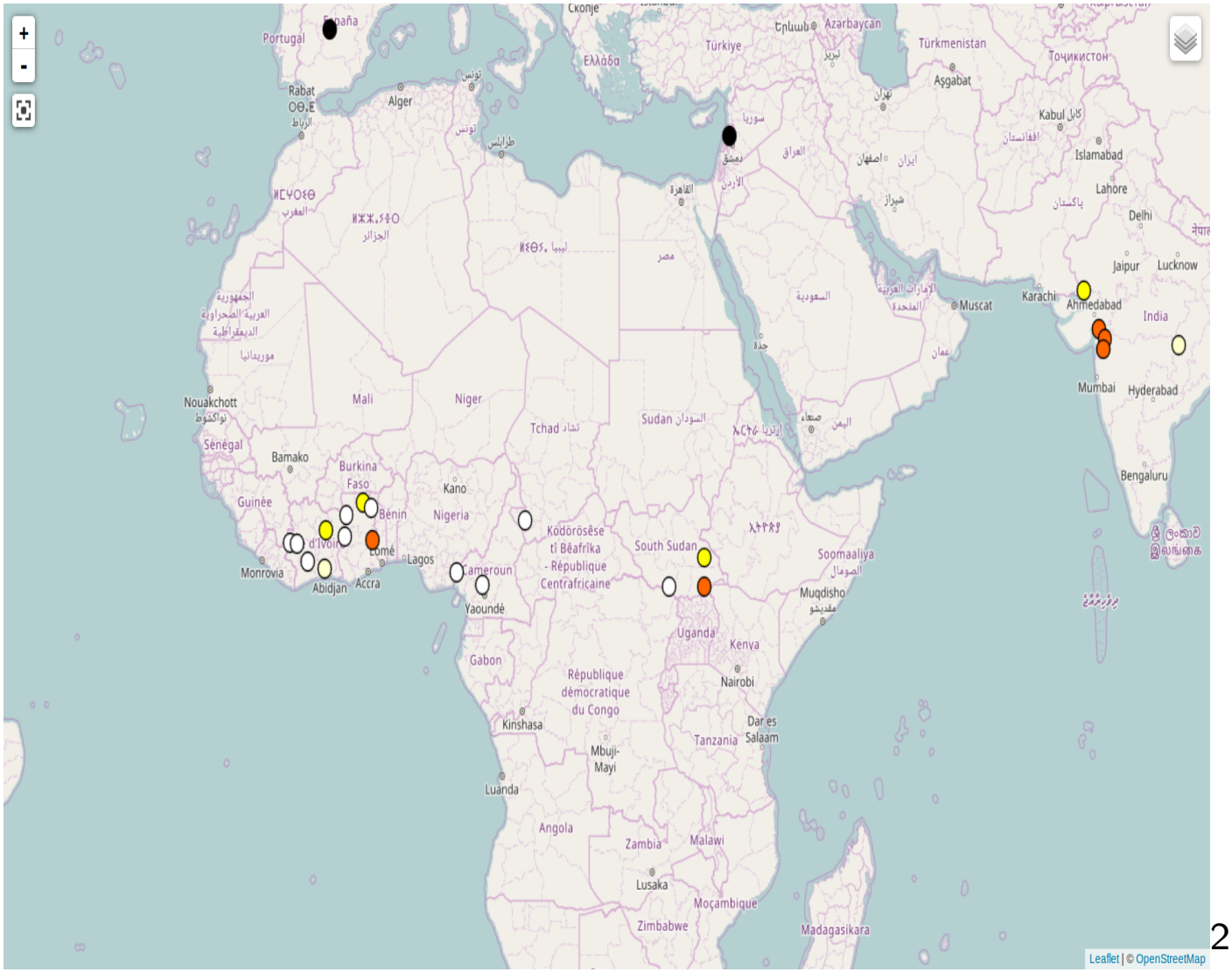
Data Points



[1]

[1] http://wals.info/feature/133A#2/-26.1/130.4

Data Points



[2]

[2] http://wals.info/feature/133A#2/-26.1/130.4

Data Points

[3]

[3] http://wals.info/feature/133A#2/-26.1/130.4

Data Points

[4]

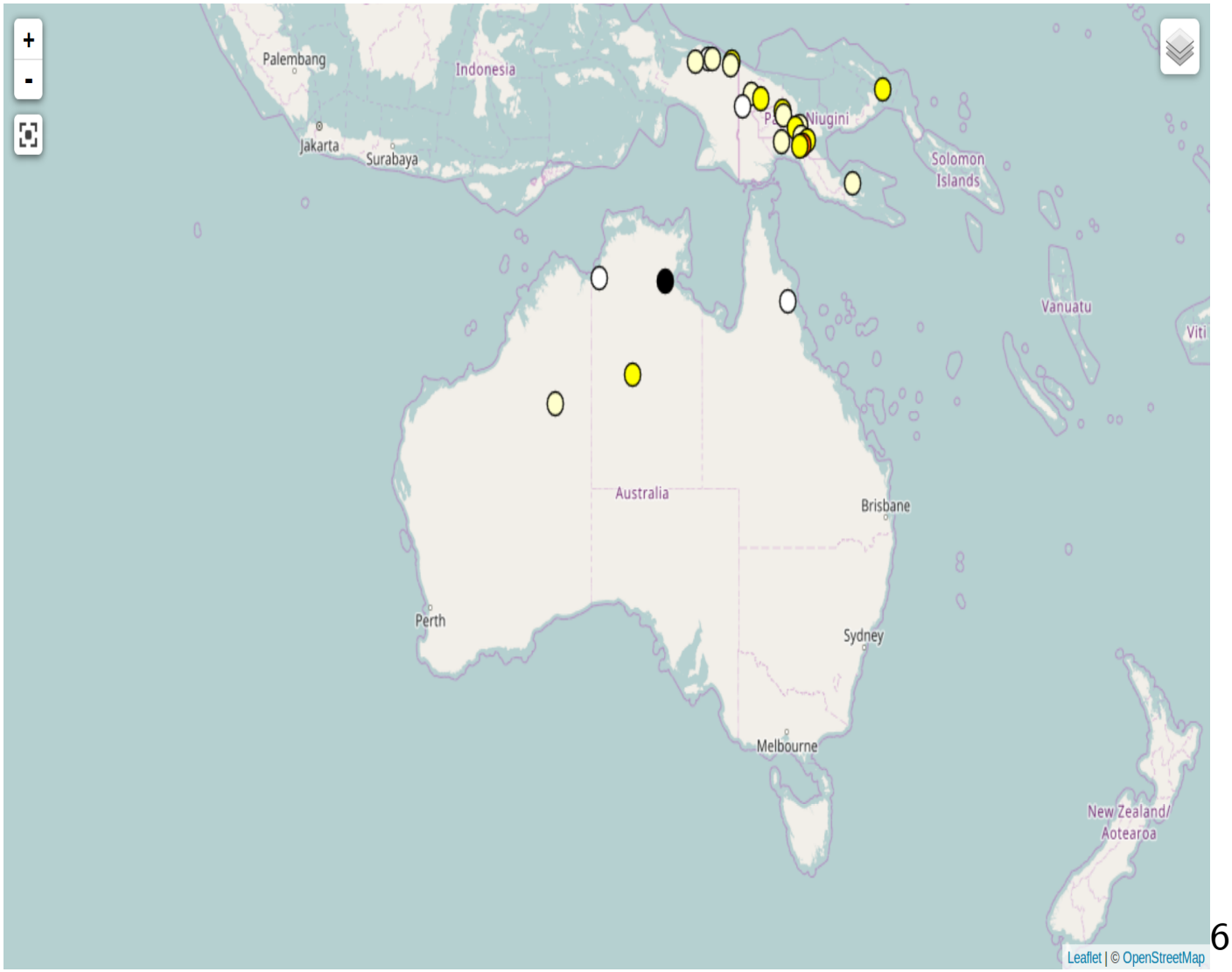
[4] http://wals.info/feature/133A#2/-26.1/130.4

Data Points

[5]

[5] http://wals.info/feature/133A#2/-26.1/130.4

Data Points



[6]

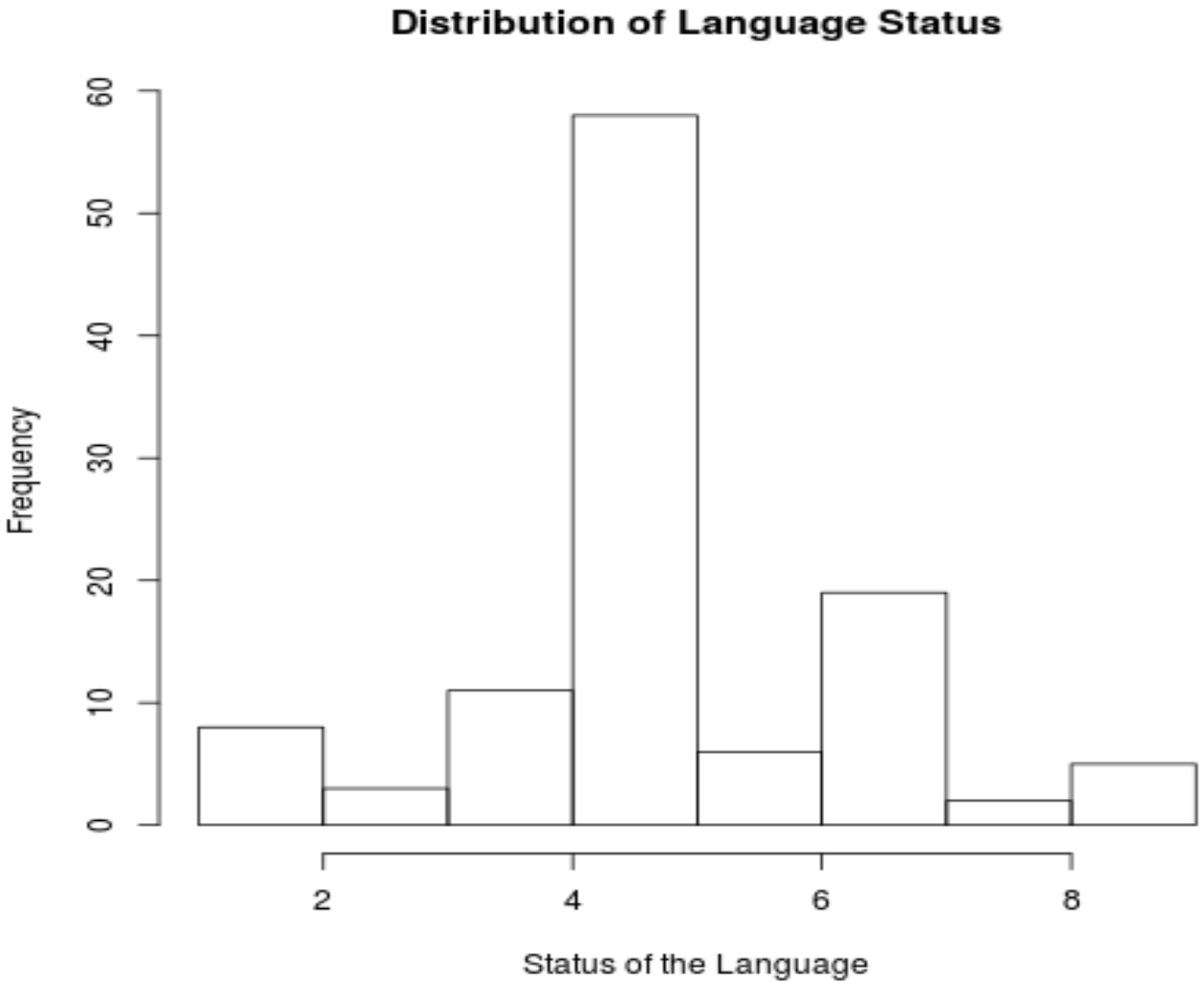
[6] http://wals.info/feature/133A#2/-26.1/130.4

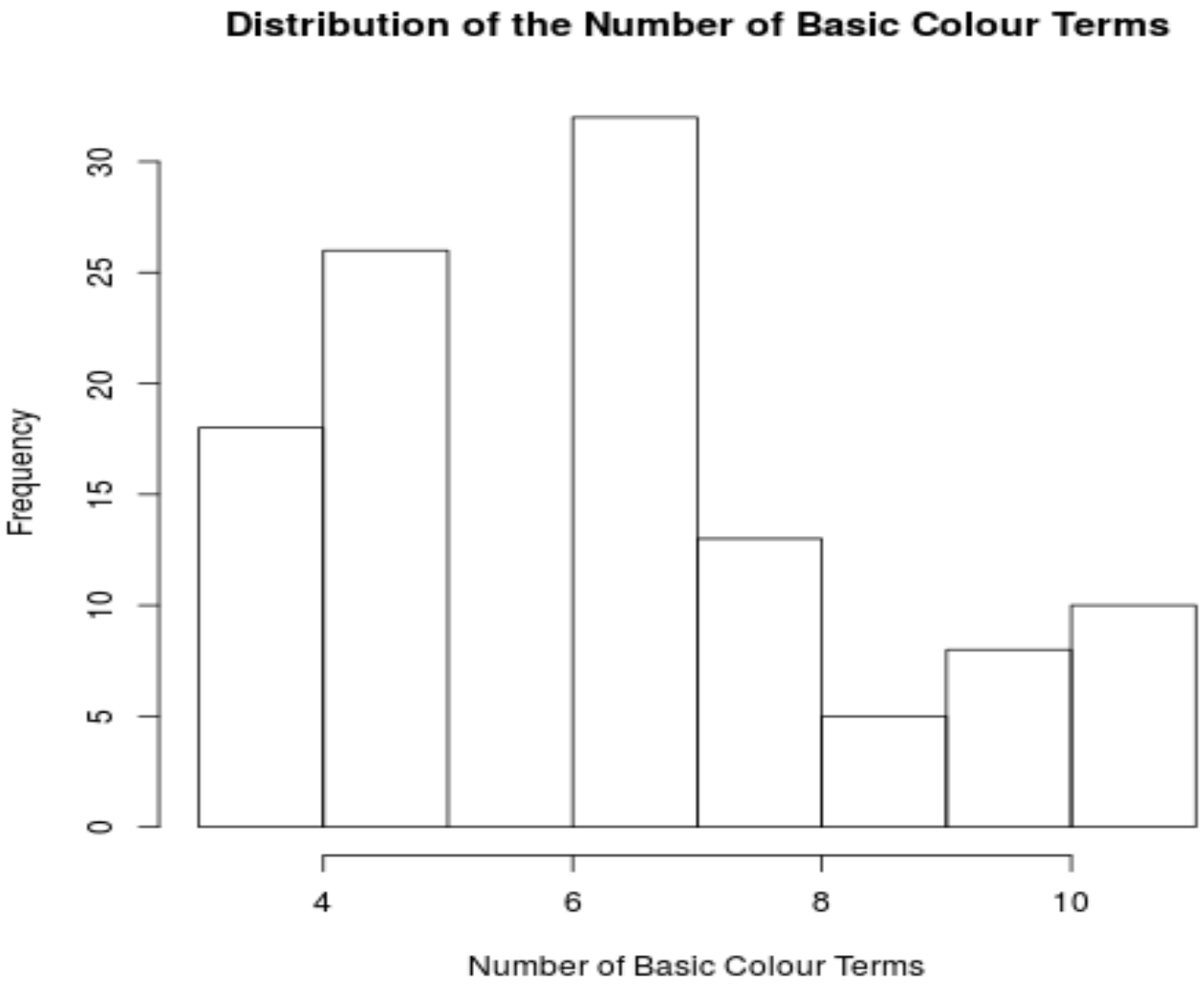
Data-Correction, Manipulation and Controversies

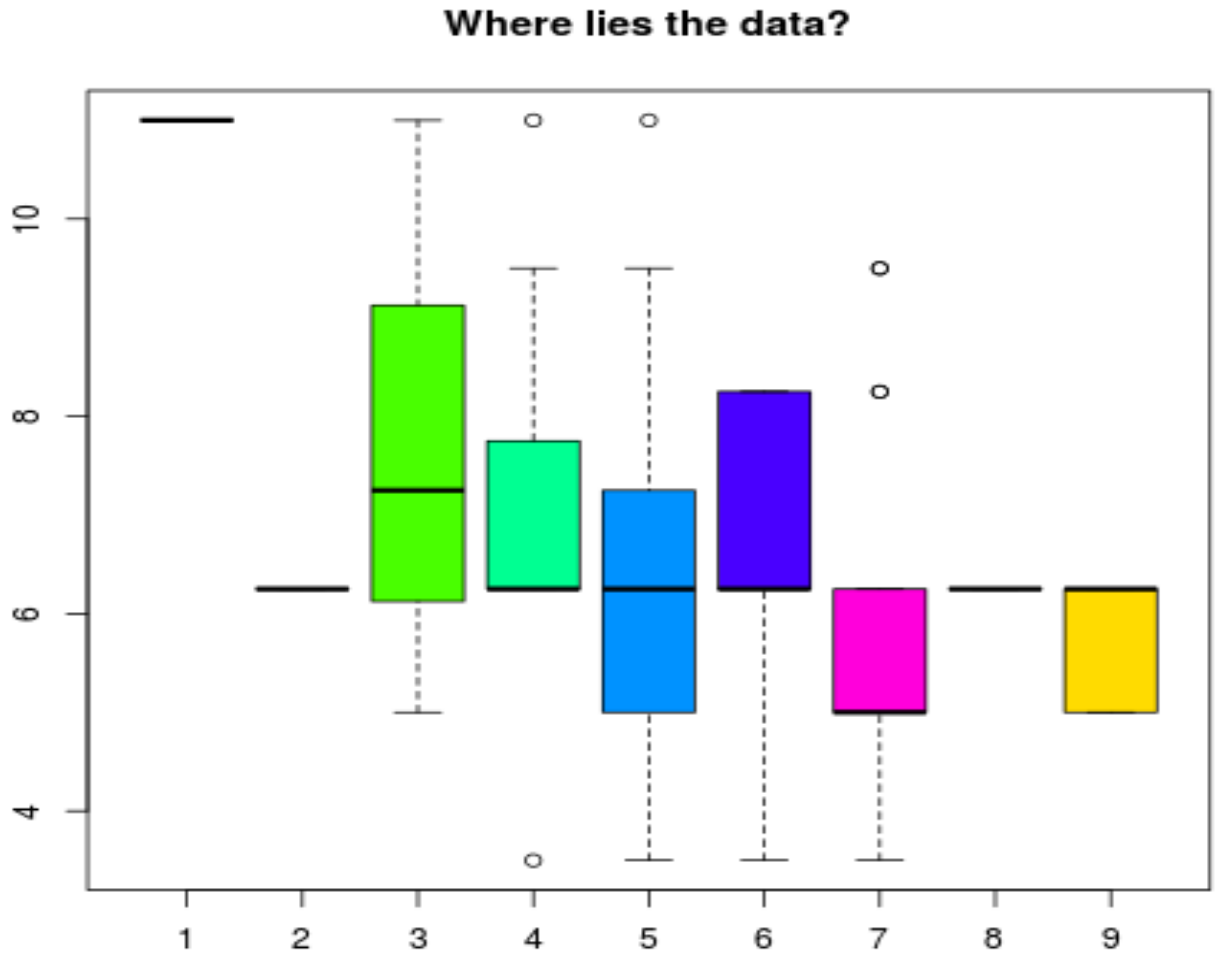
- Reassigned new values to the Ethnologue Scale
- Took the mean value for the number of basic colour terms
- Exclusion of the following data-points:
  1. Cakchiquel
  2. Campa (Axininca)
  3. Cofán
  4. Slave
  5. Aguacatec
  6. Tlapanec

- Russian is annotated with having 11 basic colour terms but it actually has 12 (two terms for blue)
- Hungarian, which is not in the sample, also has 12 (two terms for red)
- In that case it is not clear whether the both terms for red are basic colour terms

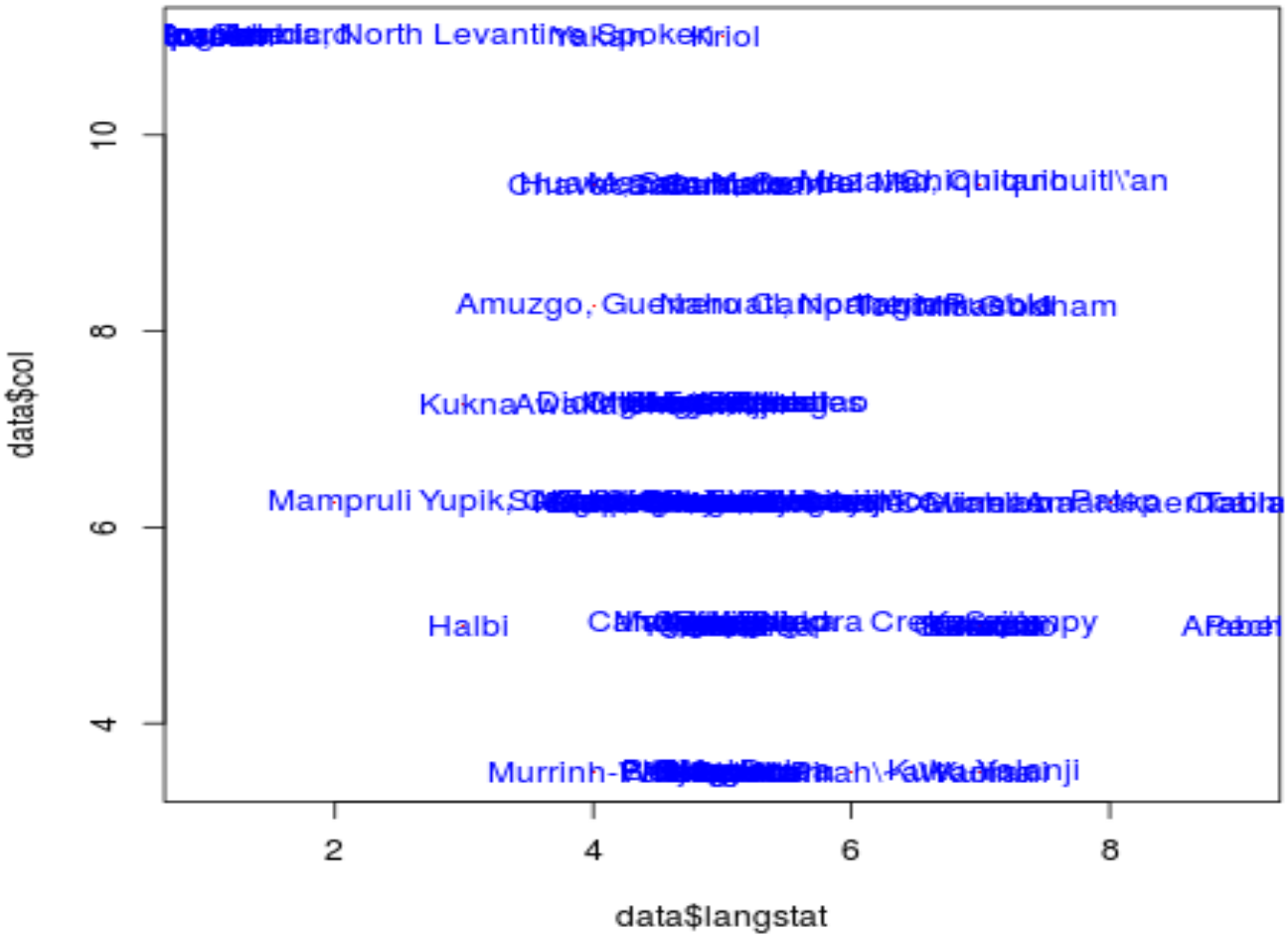
Distribution of Raw Data



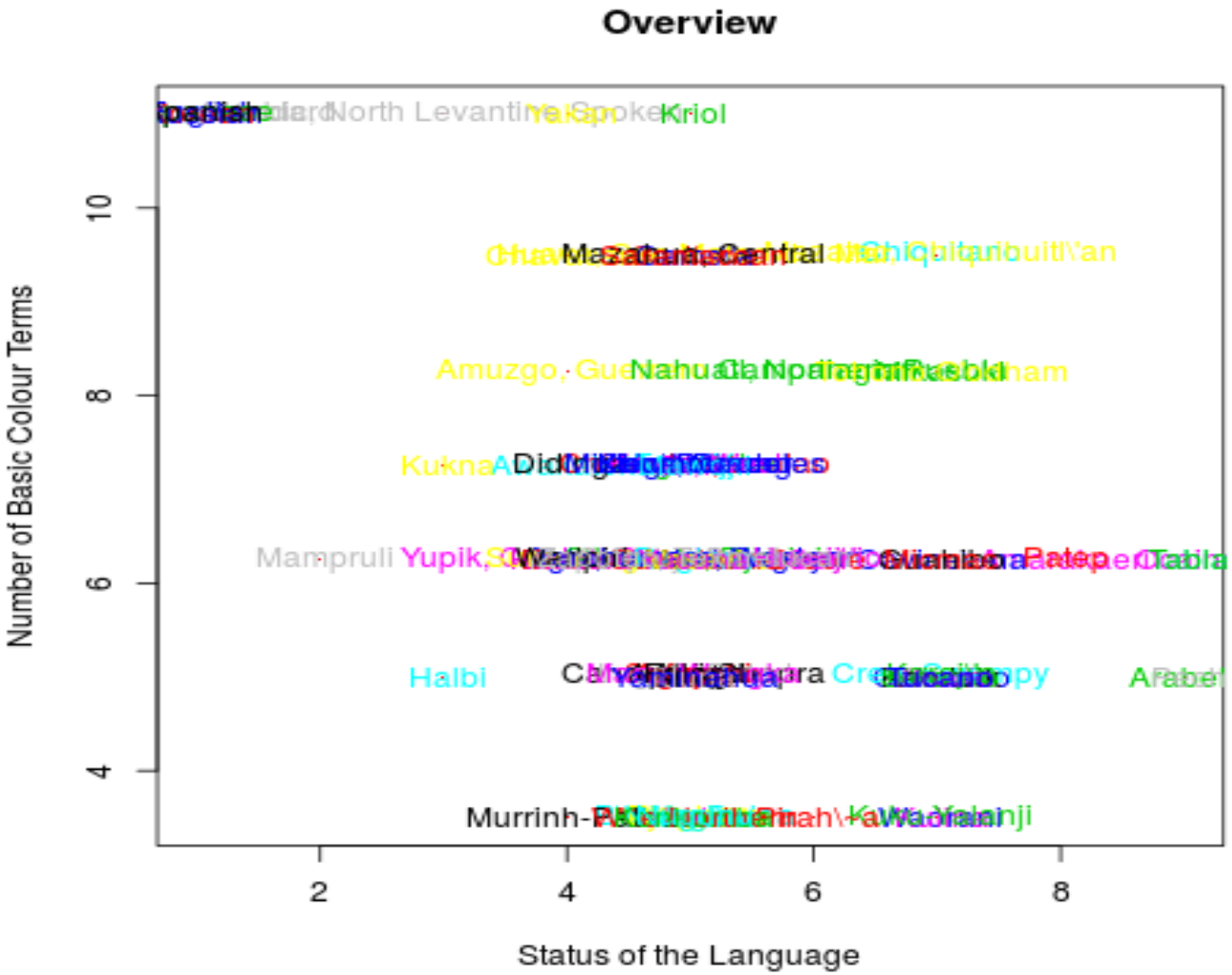




Data Distribution and Correlation



Data Distribution and Correlation



Statistical Test II

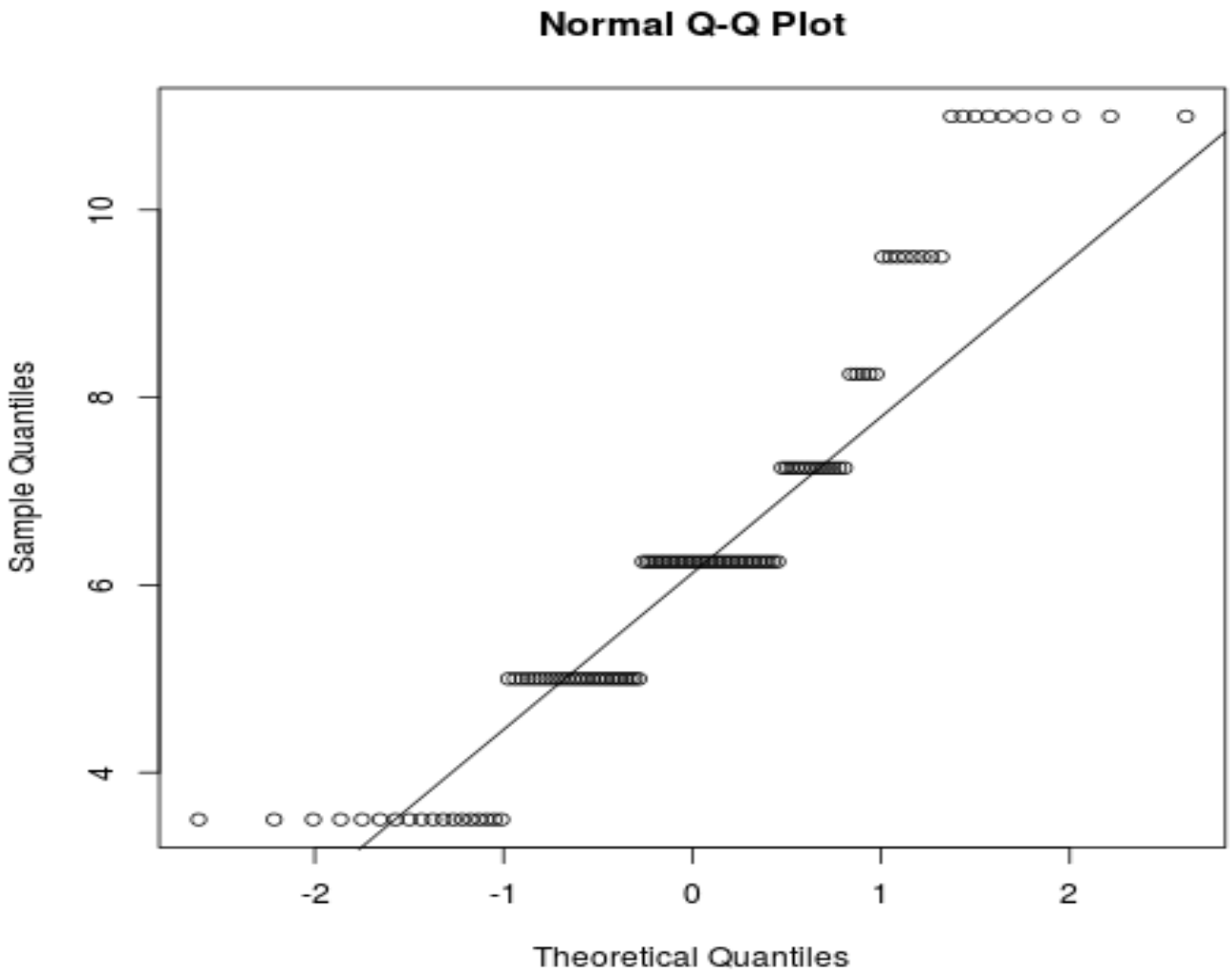
- Spearman's rank correlation rho
- performed on the two colums data$langstat and data$col with the following results:

  $S = 309583.6$, *p-value = 0.0005304*

  Correlation-rate: -0.3222381

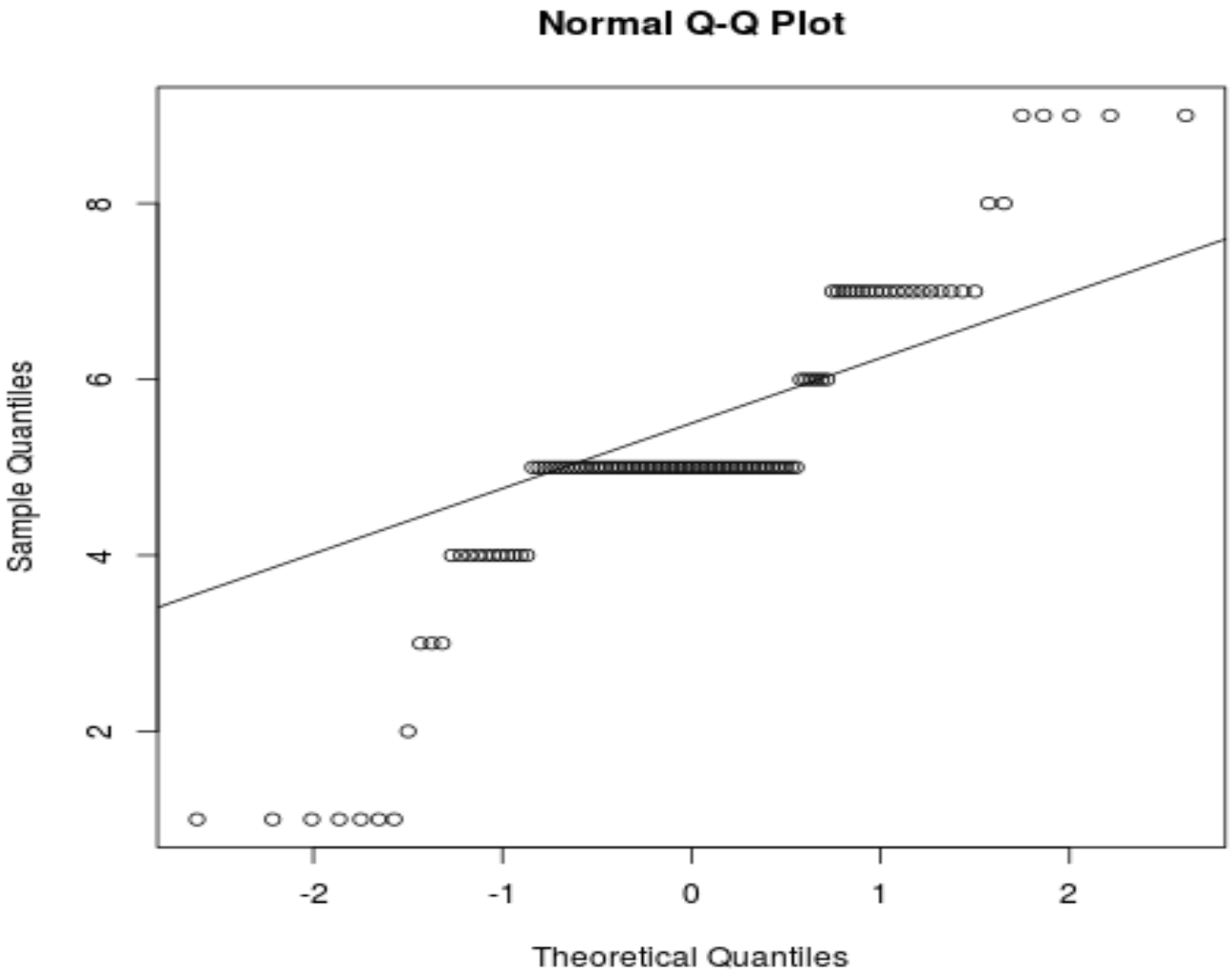
Explanations and Assessment

- **How normally distributed is the data?**



Explanations and Assessment

- **How normally distributed is the data?**



Explanation and Assessment

- Language depth as indicator for number of colour terms and different stages?
- Evolution of language does not contribute to loss of colour terms according to Berlin and Kay (1969) but maybe this has to be reassessed
- Analysis could go either way with more data as the result is not really strong
- Use more Data? Would that help?

Summary

- Negative Correlation between the two parameters
- Low p-value as indicator of significance value between the two parameters
- Results are not very strong but show a tendency - maybe a lack of data?
- Data is very biased towards minority languages - include more data

Open Research Questions

- Correlation between Status and Blue/Green distinction?
- Correlation between Status and Red/Yellow distinction?
- Correlation of Blue/Green distinction and Red/Yellow distinction?
- Correlation between the Population Size and the number of basic colour terms?
- Correlation between the Population Size and the status of the language?
- Look at specific Language Families? How do they behave? Would that even make sense?
- Any further ideas?

References I

Wals chapter: Basic colour terms.

`http://wals.info/chapter/133`.

Accessed: 2017-06-27.

Brent Berlin and Paul Kay.

Basic color terms: Their universality and evolution.

Univ of California Press, 1991.

Richard S Cook, Paul Kay, and Terry Regier.

The world color survey database.

Handbook of categorization in cognitive science, pages 223–241, 2005.

Paul Kay and Luisa Maffi.

Color appearance and the emergence and evolution of basic color lexicons.

American Anthropologist, 101:743–760, 1999.

Paul Kay and Luisa Maffi.

Green and Blue.

Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013.

References II

Paul Kay and Luisa Maffi.

Number of Basic Colour Categories.

Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013.

Paul Kay and Luisa Maffi.

Number of Non-Derived Basic Colour Categories.

Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013.

Paul Kay and Luisa Maffi.

Red and Yellow.

Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013.

Stephen C Levinson.

Yélî dnye and the theory of basic color terms.

Journal of Linguistic Anthropology, 10(1):3–55, 2000.

Vittorio Loreto, Animesh Mukherjee, and Francesca Tria.

On the origin of the hierarchy of color names.

Proceedings of the National Academy of Sciences, 109(18):6819–6824, 2012.

References III

John A Lucy.

15 the linguistics of" color.

Color categories in thought and language, page 320, 1997.