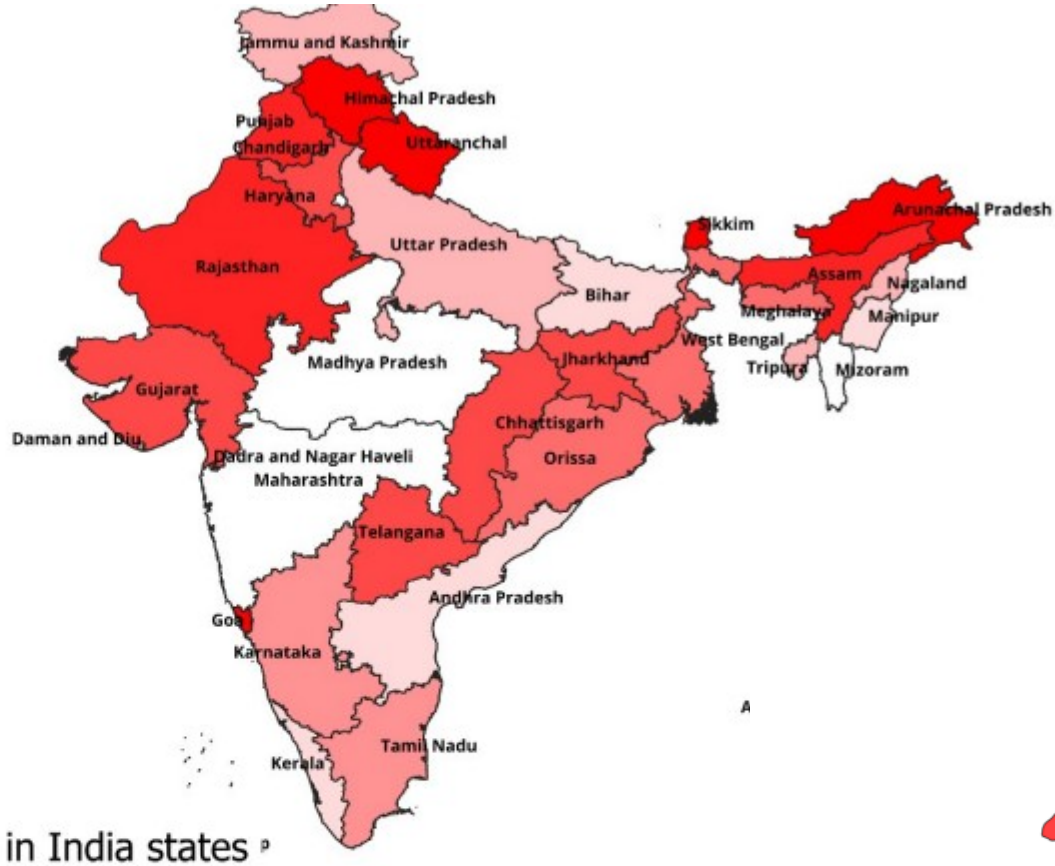
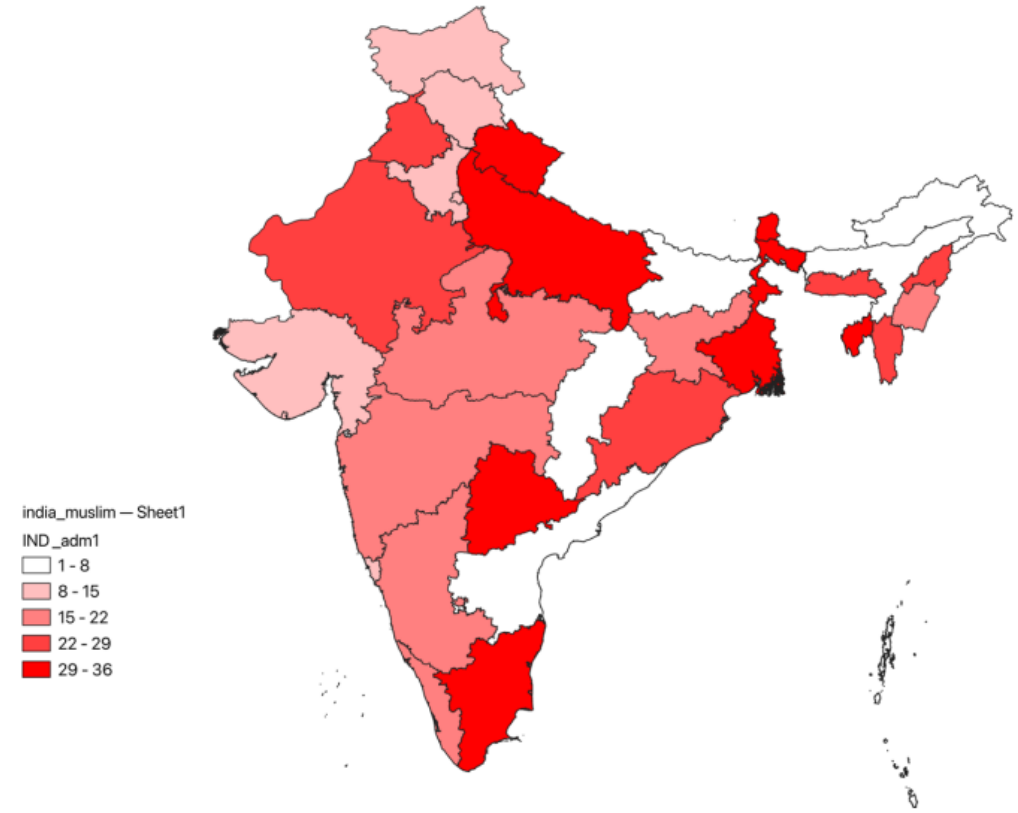


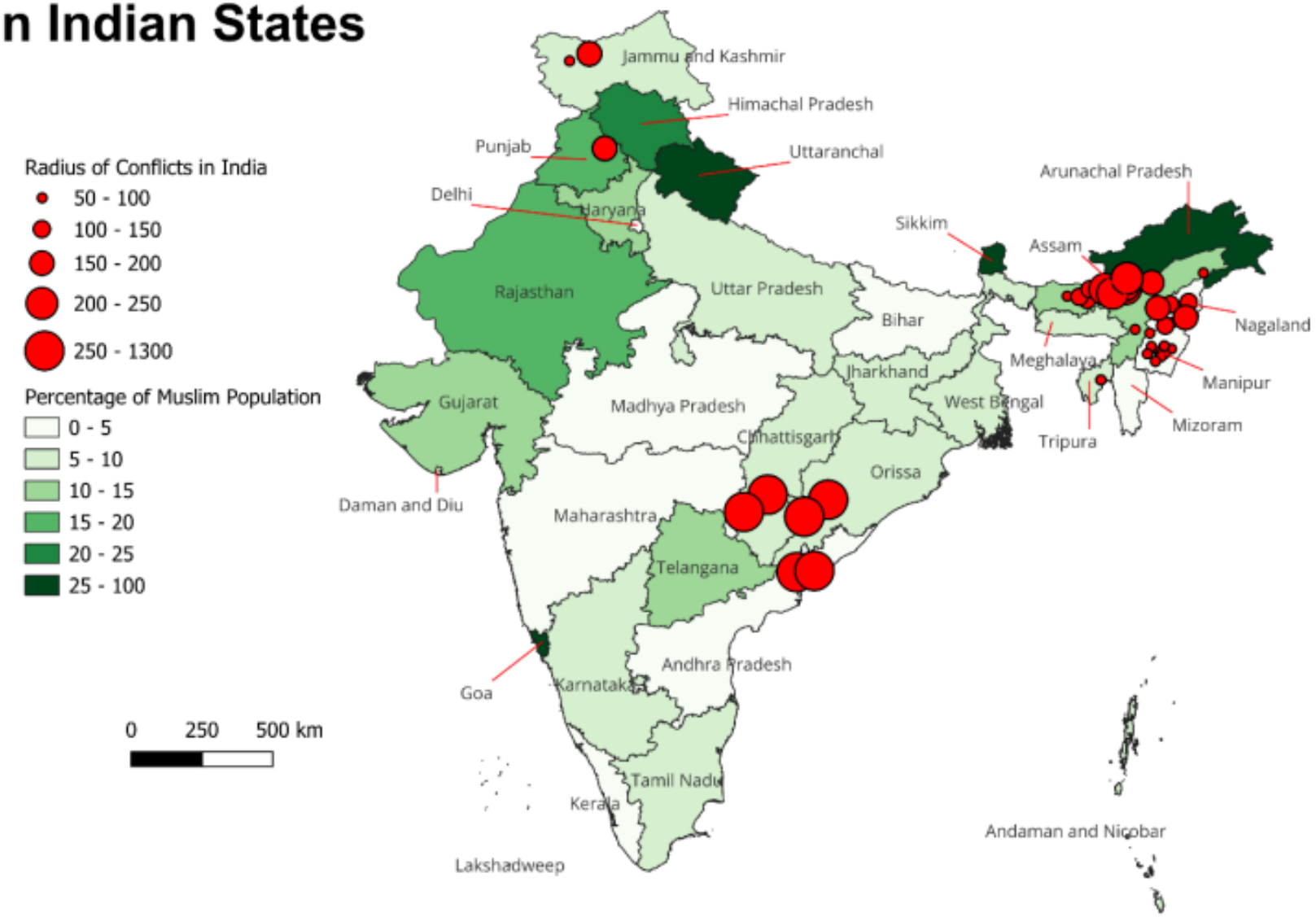

Distribution of muslim population in India states

Indian Statistical Institute

- 1.1 - 1.8
- 1.8 - 4.5
- 4.5 - 7.2
- 7.2 - 8.4
- 8.4 - 9.9
- 9.9 - 13.5
- 13.5 - 21.2
- 21.2 - 100



# Distribution of Armed Conflicts and Muslim Population in Indian States



# Data visualization – Text

# The first step

- Text has to be turned into data
- Quantitative coding
  - Human coding
  - „automatic“ (machine processed) coding
    - Almost always inclde some decision of human
    - Frequency analysis (Wordclouds)
    - Sentiment analysis - Is the text positive or negative?
    - Topic models - what topics are in the text(s)
    - Analysis of readability (how complex the text is?)
    - Analysis of „positions“

- Allocating recording units to substantial categories
  - Classifying each coded unit of text from the sample according to the category scheme
    - Coding scheme
      - Before coding (deductive)
      - During coding (inductive)
    - Unit: page, paragraph, sentence, quasisentence

# Quasi-sentence

- an argument or phrase which is the verbal expression of one idea or issue
- One sentence may include more ideas, one idea may be divided into more sentences.

- I am going to buy bread, milk and apple.

- I am going to buy bread.
- I am going to buy milk.
- I am going to buy apple.

# Text elements

- Paragraphs
  - Sentences
    - Quasi-sentences
      - N-grams
        - Words
          - Lemma
          - token

# Problems

- Language
- Tenses, adjective, male/female
- Common words (stopwords)
- Reliability of human coding
- Different understaning by different coders
- Noise in automatic coding

# Wordcloud

(c) Bush

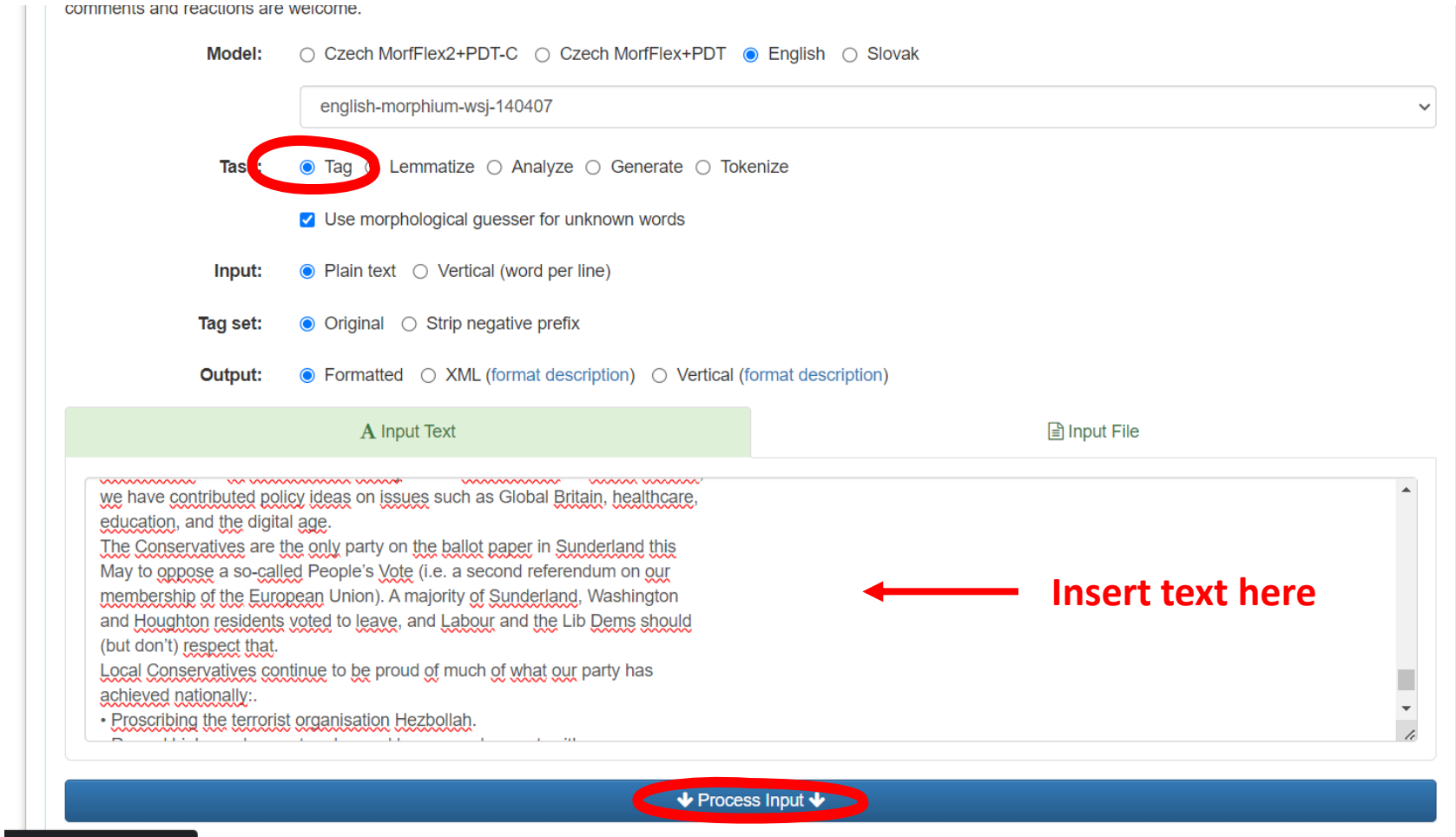
(d) Obama

# World cloud procedure

- Find a text
- Do lemmatization

# lemmatization

- https://lindat.mff.cuni.cz/services/morphodita/

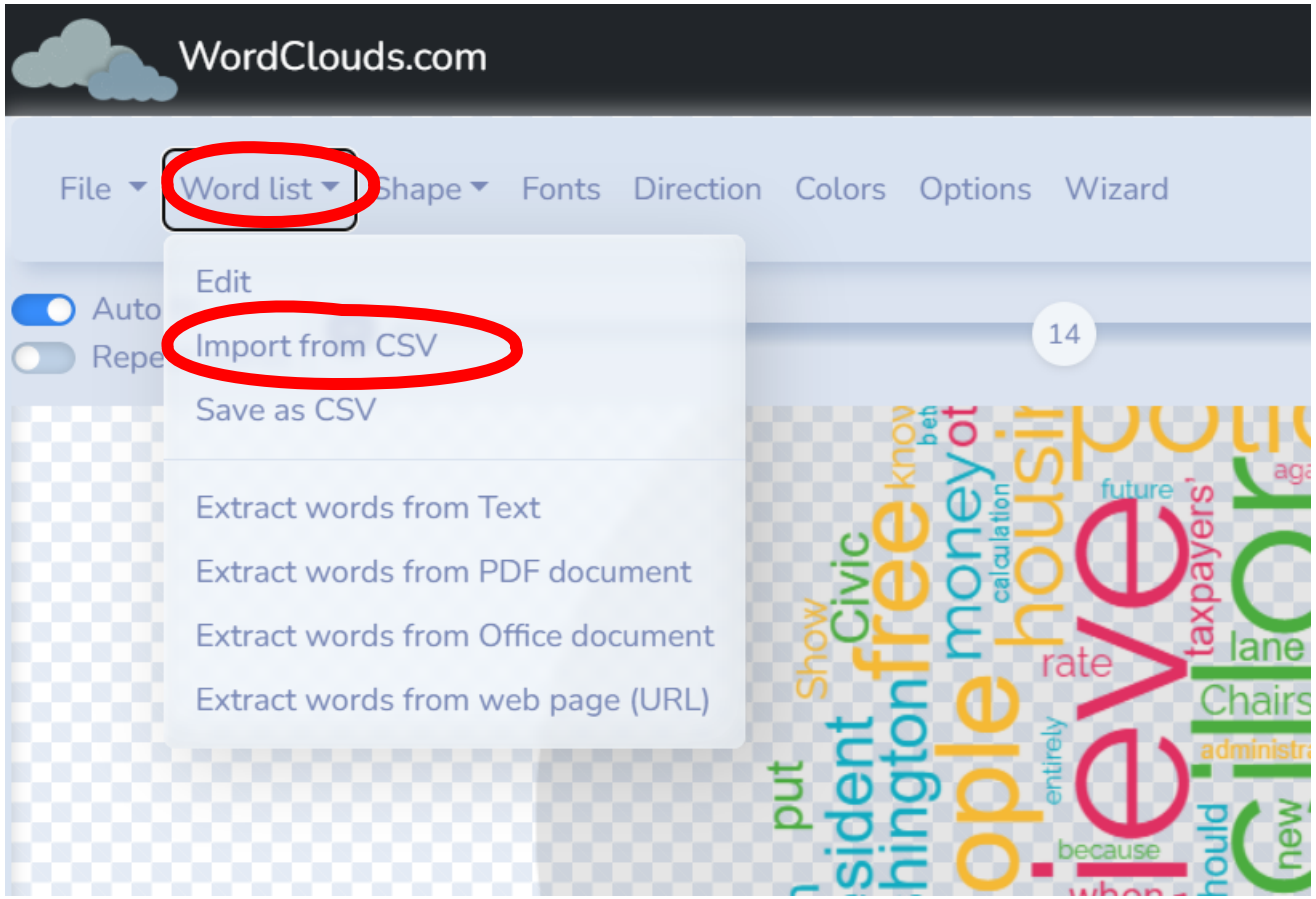


# Count and clean in excel

- Copy table „output“ into excel (export into xlm is not working well)
- Use subtotal or pivot table to count the words
- Delete stop words
  - Words which do not make difference between clouds
- The, a, an, it, I, we, they, their, have, for, in,…

- Save as csv with names:
- weight word color url
  - Weight is count, color and url may be empty

# https://www.wordclouds.com/

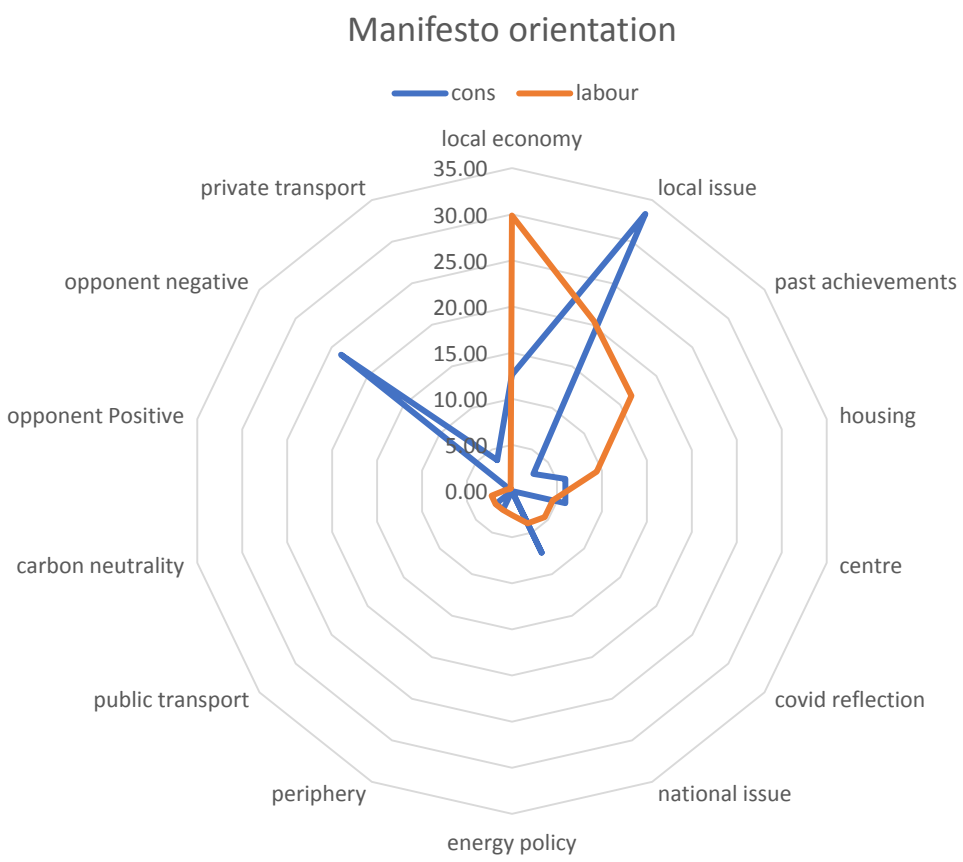


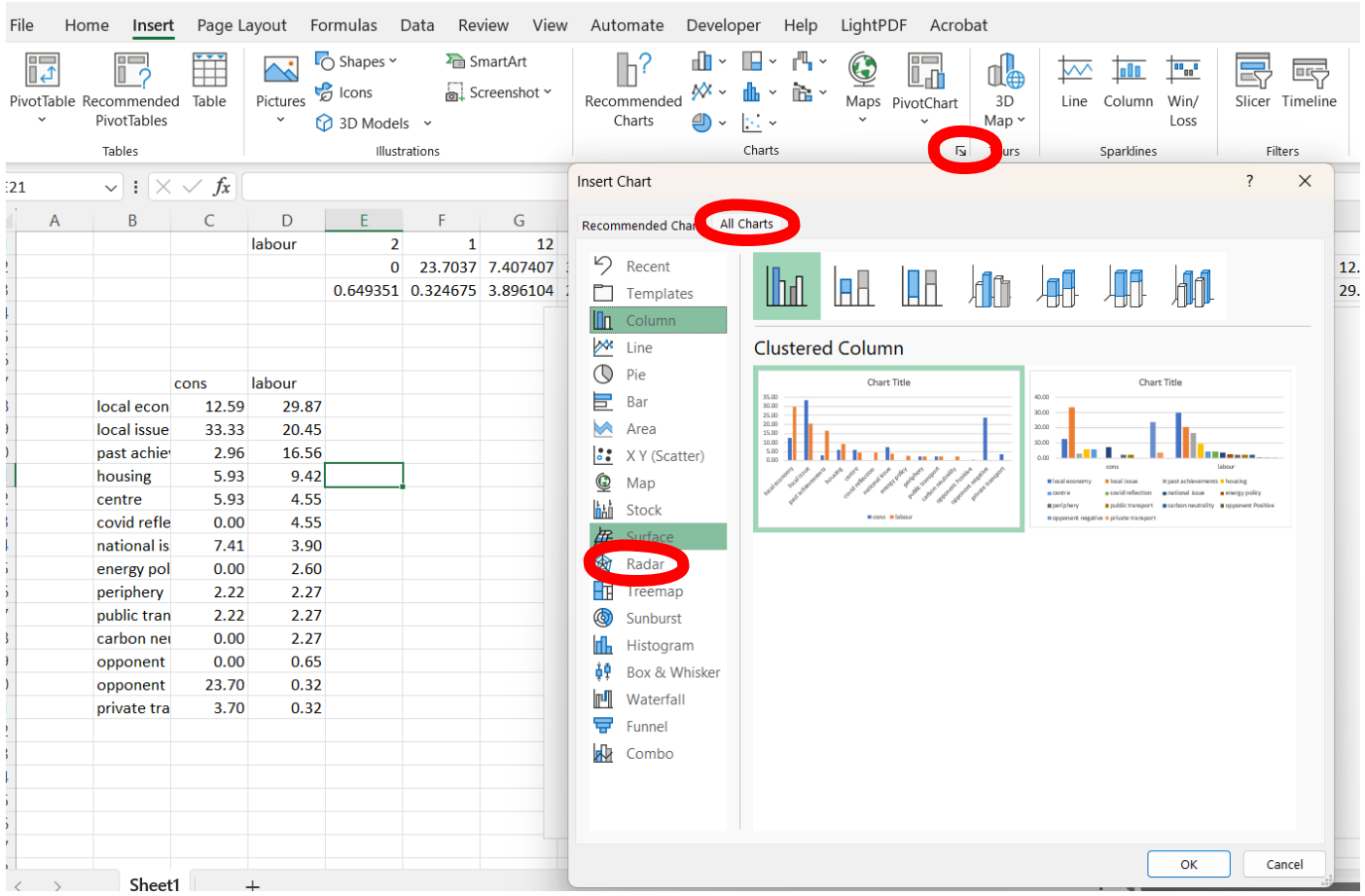
front
pay
each
decade
collection
weekly
into
only
Park
offer
opposition
Area
reduce
development
Cemetery
introduce
oppose
Community
high
end
committee
amount
addition
decision
event
cycle
policy
Centre
continue
cabinet
against
long
country
back
future
taxpayers'
elect
see
recognise
funding
vote
concert
relate
Civic
free
money
housing
other
know
between
calculation
attractive
professional
attract
fly
bin
dog
domestic
claim
Show
rate
lane
Chairs
one
duty
follow
administration
around
Girls
leadership
put
resident
Washington
people
entirely
because
new
should
gagging
make
all
well
m
db
Air
car
when
parking
which
issue
across
will
site
cut
among
alternative
last
Government
abuse
hold
orders
member
help
allocate
building
scheme
much
business
bring
explore
behaviour
meal
Coast
no
own
greenbelt
disclose
seek
empty
per
can
magazine
party
feel
Lib
move
health
Dems
year
record
award
allowance
position
provide
scrutiny
highlight
executive
litter
recent
enforcement
face
fine
time
include
good

addition
become
step
plan
city
service
opportunity
but
develop
reduce
Health
over
family
resident
build
public
there
who
take
good
place
economy
government
not
employer
strategy
centre
safe
provide
community
council
education
sector
project
use
help
people
business
include
put
programme
benefit
access
school
child
expand
most
care
great
exist
skill
key
street
Park
impact
them
home
area
other
worker
G
building
space
development
Local
bring
part
manifesto
green
carbon
work
consultation
commitment
approach
digital
activity
year
look
where
it
action
deliver
food
increase
create
can
encourage
focus
follow
do
promote
all
healthy
energy
objective
additional
enable
continue
Sunderland
housing
free
those
link
Partnership
Partner
Technology
further
make
Social
Labour
within
investment
up
now
improve
need
financial

- The cloud can be made from text without processing in excel
- Low control over stopwords

- Do not use crazy colors, shapes and text orientations
- Use the same font for all words


Manifesto orientation
cons
labour
local economy
local issue
past achievements
housing
centre
covid reflection
national issue
energy policy
periphery
public transport
carbon neutrality
opponent Positive
opponent negative
private transport
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00


File Home Insert Page Layout Formulas Data Review View Automate Developer Help LightPDF Acrobat
PivotTable Recommended PivotTables Table
Tables
Pictures Shapes Icons 3D Models SmartArt Screenshot
Illustrations
Recommended Charts Maps PivotChart
Charts
3D Map
Line Column Win/ Loss
Sparklines
Slicer Timeline
Filters
Insert Chart
Recommended Charts All Charts
Recent Templates Column Line Pie Bar Area X Y (Scatter) Map Stock Surface Radar Treemap Sunburst Histogram Box & Whisker Waterfall Funnel Combo
Clustered Column
Chart Title
OK Cancel
Sheet1