Learning R

- Lynda – Up and running with R – one of the best and 'softest' intros
- Lynda – R statistics essential training
- Code School – a very soft introduction to R
- DataCamp – Learn R for data science
- edX – Intro to R for data science – powered by Microsoft
- Coursera – Data science specialization (Johns Hopkins University) – the most time consuming option

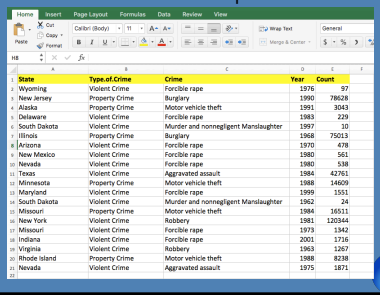
7

Resources

- QuickR
- Rbloggers – R news and tutorials (580 contributors!!)
- STHDA – Just a ton of cool stuff about stats and R
- StackOverflow – This is where you ask a question if you have a problem!☺
- Rstudio - webinars by Rstudio
- Books: The R Cookbook (Teetor, 2011), Elegant Graphics for Data Analysis (Wickham, 2016), R for Excel Analysts (Taveras, 2016), R for Data Science (Wickham & Grolemund, 2017)

8

"Traditional" excel spread-sheet



State	Type.of.Crime	Crime	Year	Count
Wyoming	Violent Crime	Forcible rape	1976	97
New Jersey	Property Crime	Burglary	1990	78628
Alaska	Property Crime	Motor vehicle theft	1991	3043
Delaware	Violent Crime	Forcible rape	1983	229
South Dakota	Violent Crime	Murder and nonnegligent Manslaughter	1997	10
Illinois	Property Crime	Burglary	1968	75013
Arizona	Violent Crime	Forcible rape	1970	478
New Mexico	Violent Crime	Forcible rape	1980	561
Nevada	Violent Crime	Forcible rape	1980	538
Texas	Violent Crime	Aggravated assault	1984	42761
Minnesota	Property Crime	Motor vehicle theft	1988	14609
Maryland	Violent Crime	Forcible rape	1999	1551
South Dakota	Violent Crime	Murder and nonnegligent Manslaughter	1962	24
Missouri	Property Crime	Motor vehicle theft	1984	16511
New York	Violent Crime	Robbery	1981	120344
Missouri	Violent Crime	Forcible rape	1973	1342
Indiana	Violent Crime	Forcible rape	2001	1716
Virginia	Violent Crime	Robbery	1963	1267
Rhode Island	Property Crime	Motor vehicle theft	1988	8238
Nevada	Violent Crime	Aggravated assault	1975	1871

How R structures data

- The data frame (observations and variables)
- Tidy data paradigm (Wickham, 2014)
- > One row per observation, one column per variable (long format)
- Also
  - Vectors (a sequence of data elements of the same basic type)
  - and matrices (rows and columns)

10

Tidy data paradigm

Variable

Observation

State	Type.of.Crime	Crime	Year	Count
Wyoming	Violent Crime	Forcible rape	1976	97
New Jersey	Property Crime	Burglary	1990	78628
Alaska	Property Crime	Motor vehicle theft	1991	3043
Delaware	Violent Crime	Forcible rape	1983	229
South Dakota	Violent Crime	Murder and nonnegligent Manslaughter	1997	10
Illinois	Property Crime	Burglary	1968	75013
Arizona	Violent Crime	Forcible rape	1970	478
New Mexico	Violent Crime	Forcible rape	1980	561
Nevada	Violent Crime	Forcible rape	1980	538
Texas	Violent Crime	Aggravated assault	1984	42761
Minnesota	Property Crime	Motor vehicle theft	1988	14609
Maryland	Violent Crime	Forcible rape	1999	1551
South Dakota	Violent Crime	Murder and nonnegligent Manslaughter	1962	24
Missouri	Property Crime	Motor vehicle theft	1984	16511
New York	Violent Crime	Robbery	1981	120344
Missouri	Violent Crime	Forcible rape	1973	1342
Indiana	Violent Crime	Forcible rape	2001	1716
Virginia	Violent Crime	Robbery	1963	1267
Rhode Island	Property Crime	Motor vehicle theft	1988	8238
Nevada	Violent Crime	Aggravated assault	1975	1871

Types of variables

- Categorical variables
  - Also known as discrete or qualitative variables (nominal, dichotomous, ordinal)
  - E.g. gender, occupation, 0/1
- Continuous variables
  - Also known as quantitative variables (interval, ratio)
  - E.g. weight, height, age, income
- Some variables could be considered in either way.
  - E.g. Likert-scale items: attractiveness rating on 5-point scale continuous or categorical (5 levels)

12

Types of variables cont'

Categorical variable

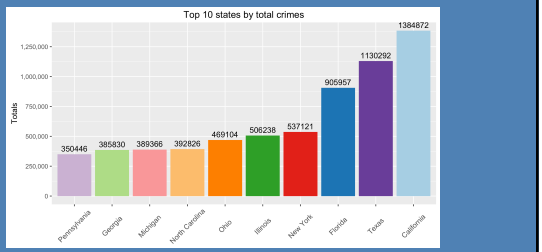
Continuous variable

State	Type.of.Crime	Crime	Year	Count
Wyoming	Violent Crime	Forcible rape	1976	97
New Jersey	Property Crime	Burglary	1990	78628
Alaska	Property Crime	Motor vehicle theft	1991	3043
Delaware	Violent Crime	Forcible rape	1983	229
South Dakota	Violent Crime	Murder and nonnegligent Manslaughter	1997	10
Illinois	Property Crime	Burglary	1968	75013
Arizona	Violent Crime	Forcible rape	1970	478
New Mexico	Violent Crime	Forcible rape	1980	561
Nevada	Violent Crime	Forcible rape	1980	538
Texas	Violent Crime	Aggravated assault	1984	42761
Minnesota	Property Crime	Motor vehicle theft	1988	14609
Maryland	Violent Crime	Forcible rape	1999	1551
South Dakota	Violent Crime	Murder and nonnegligent Manslaughter	1962	24
Missouri	Property Crime	Motor vehicle theft	1984	16511
New York	Violent Crime	Robbery	1981	120344
Missouri	Violent Crime	Forcible rape	1973	1342
Indiana	Violent Crime	Forcible rape	2001	1716
Virginia	Violent Crime	Robbery	1963	1267
Rhode Island	Property Crime	Motor vehicle theft	1988	8238
Nevada	Violent Crime	Aggravated assault	1975	1871

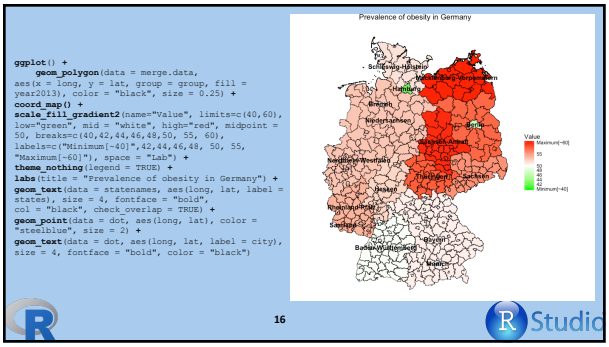
ggplot2

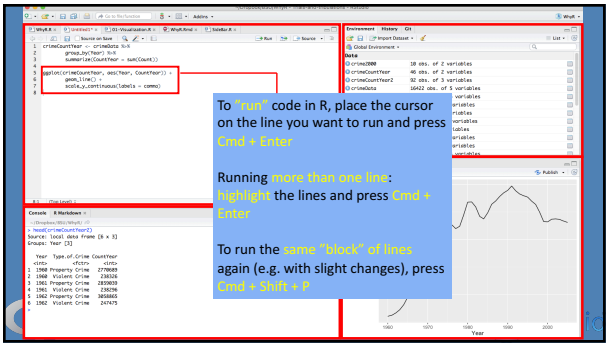
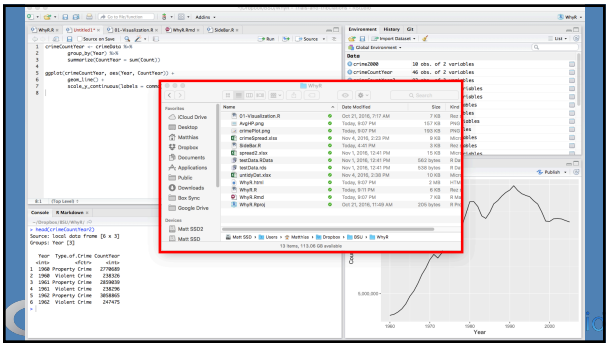
- http://docs.ggplot2.org/current/index.html
- Written by Hadley Wickham – current version 2.2.0 (2016)
- Based on the grammar of graphics (Wilkinson, 2005) – plot made up of layers

14



15


ggplot() +
geom_polygon(data = merge.data, aes(x = long, y = lat, group = group, fill = year2013), color = "black", size = 0.25) +
coord_map() +
scale_fill_gradient2(name="Value", limits=c(40,60), low="green", mid = "white", high="red", midpoint = 50, breaks=c(40,42,44,46,48,50, 55, 60), labels=c("Minimum[-40]",42,44,46,48, 50, 55, "Maximum[-60]"), space = "Lab") +
theme_nothing(legend = TRUE) +
labs(title = "Prevalence of obesity in Germany") +
geom_text(data = statenames, aes(long, lat, label = states), size = 4, fontface = "bold", col = "black", check_overlap = TRUE) +
geom_point(data = dot, aes(long, lat), color = "steelblue", size = 2) +
geom_text(data = dot, aes(long, lat, label = city), size = 4, fontface = "bold", color = "black")
Prevalence of obesity in Germany
16
RStudio


To "run" code in R, place the cursor on the line you want to run and press Cmd + Enter
Running more than one line: highlight the lines and press Cmd + Enter
To run the same "block" of lines again (e.g. with slight changes), press Cmd + Shift + P