

Инструменты для визуализации данных (и научных исследований вообще)

Алексей Кнорре

Email: aknorre@eu.spb.ru

WWW: alexeyknorre.ru

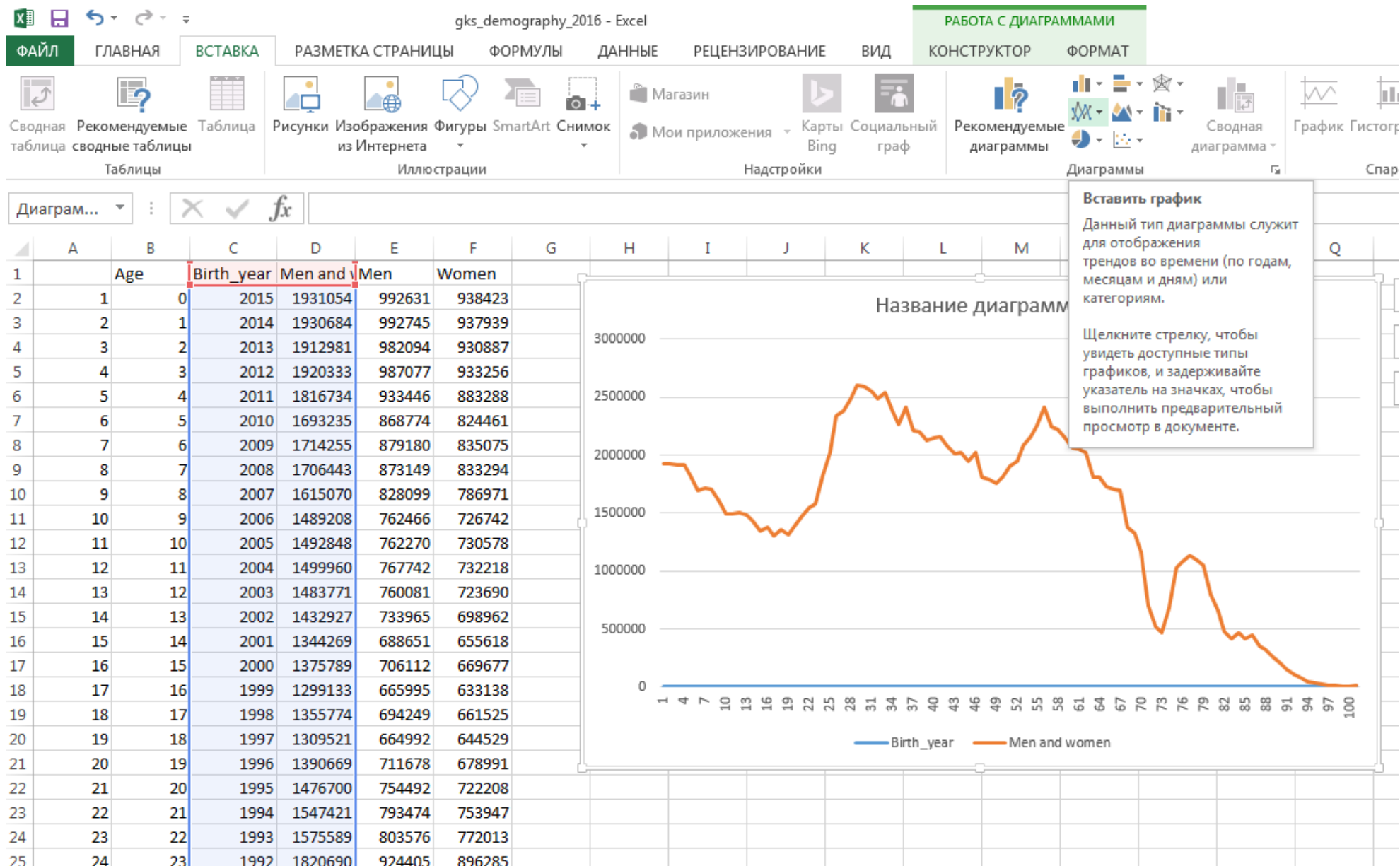
Ссылка на презентацию: alexeyknorre.ru/courses/datavis2016/datavis-5.pdf

План занятия

- Инструменты для визуализации данных
 - Excel
 - SPSS, другие статистические пакеты
 - Adobe Illustrator
- Интерактивные веб-визуализации
 - JavaScript-библиотеки (D3.js), Shiny в RStudio
- Воспроизводимость (reproducibility) исследований
 - Код и данные, системы контроля версий, публикация онлайн
- Немного R: введение и первые шаги
 - R / Rstudio. Что это?

Задача учёного – сделать своё
исследование **интересным**

Excel



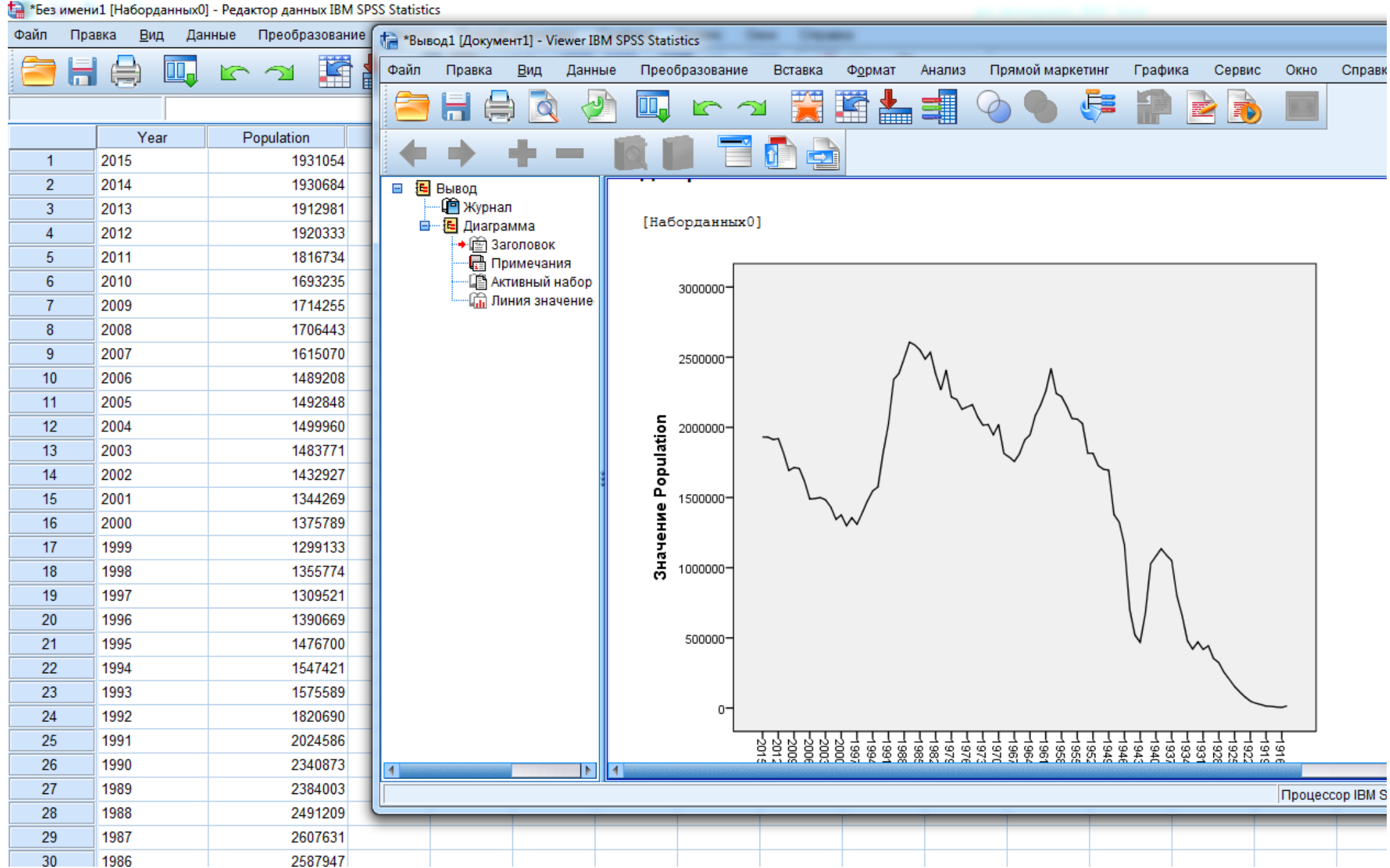
Почему использовать Excel – очень плохо?

Как экономисты получили результаты, оправдывающие политику жесткой экономии, по ним были приняты большие политические решения, а потом в исследовании нашли ошибки, связанные с обработкой данных в Excel:

Случай Reinhart and Rogoff (2011)

URL: https://en.wikipedia.org/wiki/Growth_in_a_Time_of_Debt

SPSS



Профессиональная инфографика: Adobe Illustrator

<https://www.youtube.com/watch?v=gJPnbk9L5Mw>

Интерактивные веб-визуализации: D3.js

[Overview](#) [Examples](#) [Documentation](#) [Source](#)



Примеры: <https://d3js.org/>

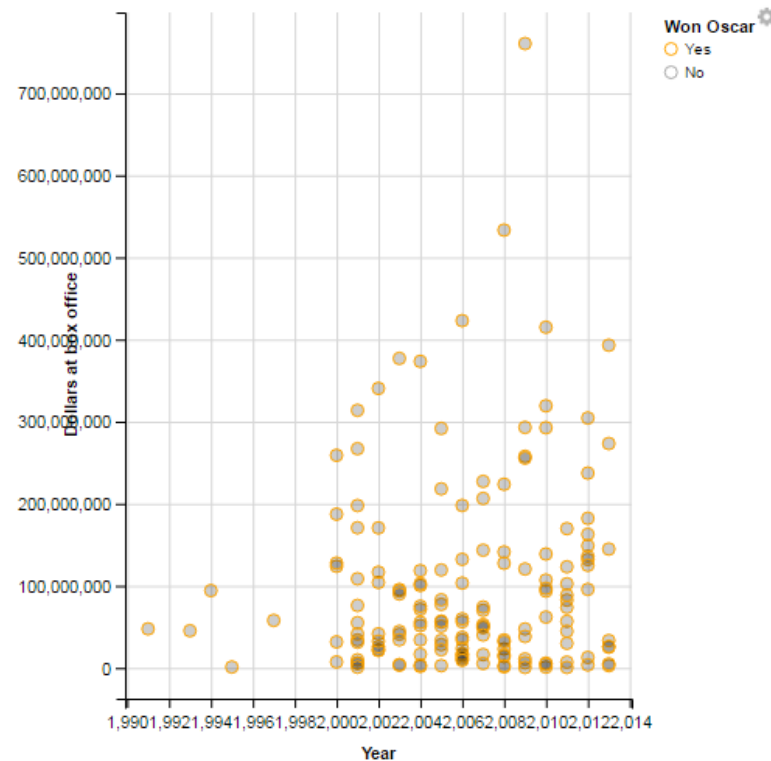
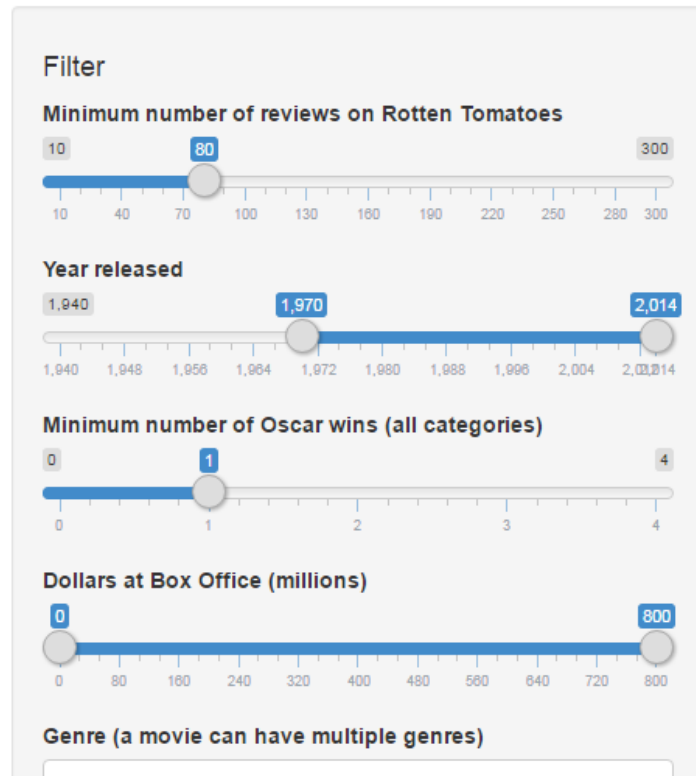
Введение: <http://square.github.io/intro-to-d3/>

Интерактивные веб-визуализации: Shiny

Shiny by RStudio

BACK TO GALLERY

Movie explorer



Визуализация фильмов: <http://shiny.rstudio.com/gallery/movie-explorer.html>

Воспроизводимые исследования (reproducible research)

Что такое воспроизводимость? (reproducible research)

Reproduction – повторный анализ исходных данных до получения такого же результата.

Replication – полный повтор всего исследования, включая сбор данных.

Зачем мне знать, что такое
воспроизводимость?



Научный метод вкратце

Проблемы науки, связанные с воспроизводимостью

- Фальсификация данных

Article Views

► Abstract

► Full Text

► Full Text (PDF)

► Figures Only

► Supplementary
Materials

Article Tools

► Save to My Folders

► Download Citation

► Alert Me When Article is
Cited

► Post to CiteULike

► Article Usage Statistics

► E-mail This Page

► Rights & Permissions

► Commercial Reprints
and E-Prints

► View PubMed Citation

Related Content

Similar Articles In:

► Science Magazine

This article has been retracted**An Expression of Concern has been published for this article**

Read Full Text to Comment (0)

< Prev | Table of Contents | Next >

Science 12 December 2014:
Vol. 346 no. 6215 pp. 1366-1369
DOI: 10.1126/science.1256151

REPORT

When contact changes minds: An experiment on transmission of support for gay equalityMichael J. LaCour¹, Donald P. Green²

± Author Affiliations

ABSTRACT

EDITOR'S SUMMARY

Can a single conversation change minds on divisive social issues, such as same-sex marriage? A randomized placebo-controlled trial assessed whether gay ($n = 22$) or straight ($n = 19$) messengers were effective at encouraging voters ($n = 972$) to support same-sex marriage and whether attitude change persisted and spread to others in voters' social networks. The results, measured by an unrelated panel survey, show that both gay and straight canvassers produced large effects initially, but only gay canvassers' effects persisted in 3-week, 6-week, and 9-month follow-ups. We also find strong evidence of within-household transmission of opinion change, but only in the wake of conversations with gay canvassers. Contact with gay canvassers further caused substantial change in the ratings of gay men and lesbians more generally. These large, persistent, and contagious effects were confirmed by a follow-up experiment. Contact with minorities coupled with discussion of issues pertinent to them is capable of producing a cascade of opinion change.

Проблемы науки, связанные с воспроизводимостью

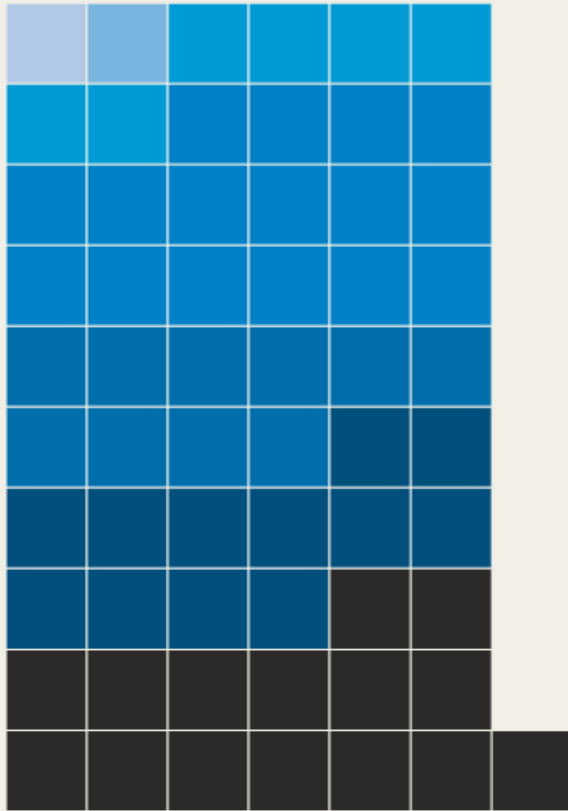
- Фальсификация данных
- Изменчивость признаков (в социальных науках)

RELIABILITY TEST

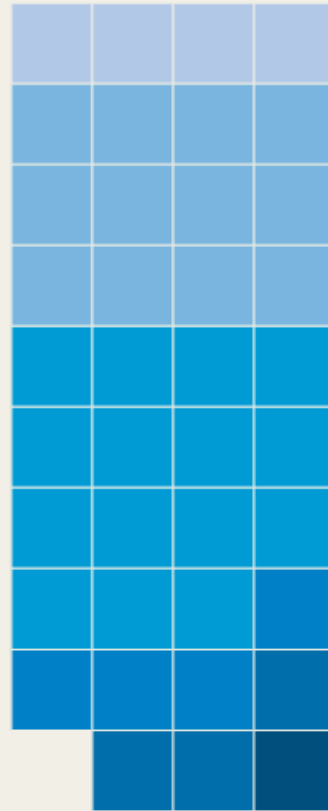
An effort to reproduce 100 psychology findings found that only 39 held up.* But some of the 61 non-replications reported similar findings to those of their original papers.

Did replicate match original's results?

NO: 61



YES: 39



Replicator's opinion: How closely did findings resemble the original study:

- Virtually identical
- Extremely similar
- Very similar
- Moderately similar
- Somewhat similar
- Slightly similar
- Not at all similar

* based on criteria set at the start of each study

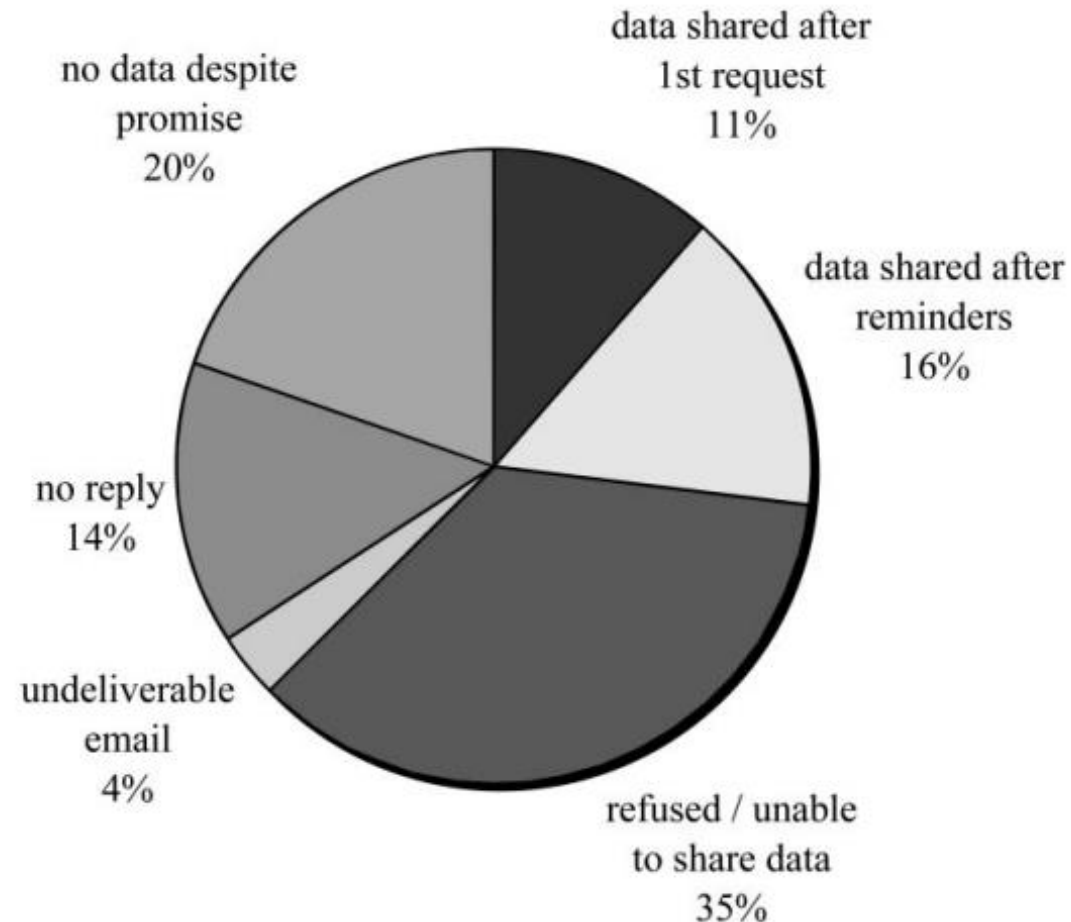
**Результаты большого
эксперимента по
воспроизводству 100 опытов в
ПСИХОЛОГИИ
(картинка с сайта nature.com)**

Проблемы науки, связанные с воспроизводимостью

- Фальсификация данных
- Изменчивость признаков (в социальных науках)
- Закрытость исследовательского процесса

Figure 1.

Percentages of Empirical Articles' Corresponding Authors in Different Response Categories



ИСТОЧНИК:

The poor availability of psychological research data for reanalysis.

Wicherts JM, Borsboom D, Kats J, Molenaar D // American Psychologist

Проблемы в науке

- Фальсификация данных
- Изменчивость признаков (в социальных науках)
- Закрытость исследовательского процесса

Как решить эти проблемы?

ВОСПРОИЗВОДИМОСТЬ



Reproducible research как способ снизить ошибки в науке

Максимальная автоматизация исследовательского процесса

- обработка и анализ (R, Python),
- генерация рисунков, табличек и отчётов (LaTeX+knitr),
- система контроля версий данных и текста (git)

Reproducible research как способ снизить ошибки в науке

Максимальная автоматизация исследовательского процесса

- обработка и анализ (R, Python),
- генерация рисунков, табличек и отчётов (LaTeX+knitr),
- система контроля версий данных и текста (git)

Открытые данные и процедуры анализа

Всё-таки, зачем мне знать о
воспроизводимости в науке?

Меньше ложных исследований,
больше доверия научным результатам.

Научная статья с открытыми данными > просто научная статья

Как делать воспроизводимые исследования?

Healy Kieran. The Plain Person's Guide to Plain Text Social Science

URL: <https://kieranhealy.org/files/papers/plain-person-text.pdf>

R / RStudio

Почему R?

Практические аргументы

- R – сам по себе язык программирования. Можно делать что угодно, даже заказать пиццу.

Источник: <http://xavier-fim.net/R/eng.html>

Практические аргументы

- R – сам по себе язык программирования. Можно делать что угодно, даже заказать пиццу.
- R – очень хорошо документированный проект. На все возникающие вопросы уже дали ответы много раз в Интернете.

Практические аргументы

- R – сам по себе язык программирования. Можно делать что угодно, даже заказать пиццу.
- R – очень хорошо документированный проект. На все возникающие вопросы уже дали ответы много раз в Интернете.
- Куча расширений, модулей и библиотек для кучи задач. Всё уже сделано за нас.

Практические аргументы

- R – сам по себе язык программирования. Можно делать что угодно, даже заказать пиццу.
- R – очень хорошо документированный проект. На все возникающие вопросы уже дали ответы много раз в Интернете.
- Куча расширений, модулей и библиотек для кучи задач. Всё уже сделано за нас.
- Открытый код. Можно понять, как оно работает в любой момент и изменить под себя.

Практические аргументы

- R – сам по себе язык программирования. Можно делать что угодно, даже заказать пиццу.
- R – очень хорошо документированный проект. На все возникающие вопросы уже дали ответы много раз в Интернете.
- Куча расширений, модулей и библиотек для кучи задач. Всё уже сделано за нас.
- Открытый код. Можно понять, как оно работает в любой момент и изменить под себя.
- R – хороший друг. Дружит почти со всеми форматами хранения данных.

Этические аргументы

- R - бесплатный и открытый. Не надо покупать/пиратить дорогие статистические пакеты вроде SPSS.

Этические аргументы

- R - бесплатный и открытый. Не надо покупать/пиратить дорогие статистические пакеты вроде SPSS.
- R работает на воспроизводимость исследований.

Этические аргументы

- R - бесплатный и открытый. Не надо покупать/пиратить дорогие статистические пакеты вроде SPSS.
- R работает на воспроизводимость исследований.
- Чем больше вы на нём работаете, публикуете и учите, тем больше это будут делать другие люди.

R / RStudio: можно заказать пиццу

Roger D. Peng wrote:

*> I'm not too concerned about your colleague's view about R. S/He doesn'
> have to like it, and I don't think anyone actually believes that R is
> designed to make *everyone* happy. For me, R does about 99% of the
> things I need to do, but sadly, when I need to order a pizza, I still
> have to pick up the telephone.*

There are several chains of pizzerias in the U.S. that provide for Internet-based ordering (e.g. www.papajohnsonline.com) so, with the Internet modules in R, it's only a matter of time before you will have a pizza-ordering function available.

Источник: <http://tolstoy.newcastle.edu.au/R/help/04/06/1672.html>

Запускаем RStudio...

Домашнее задание

Следующее занятие – мастер-класс по анализу и визуализации данных в Rstudio

- Какие данные будем использовать?
- Взять ноутбуки с установленным R/Rstudio:
R <https://cran.r-project.org/bin/windows/base/>
Rstudio <https://download1.rstudio.org/RStudio-1.0.44.exe>

Спасибо! Вопросы?

Ссылка на презентацию:

alexeyknorre.ru/courses/datavis2016/datavis-5.pdf

Алексей Кнорре

Email: aknorre@eu.spb.ru

WWW: alexeyknorre.ru