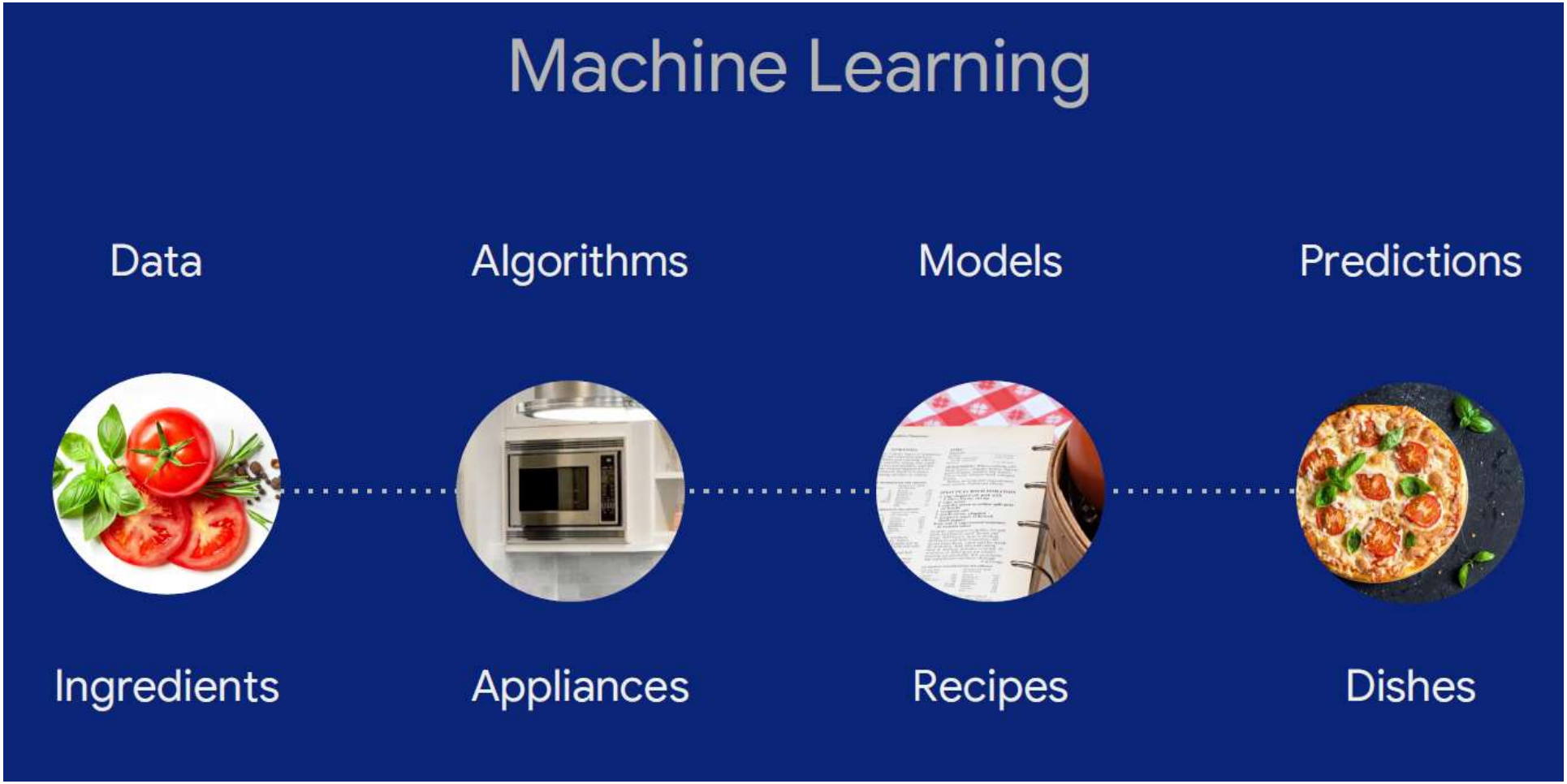
# Анализ данных с использованием языка программирования R

Минюкович Екатерина Александровна
к.э.н., доцент

miniukovich@bsu.by

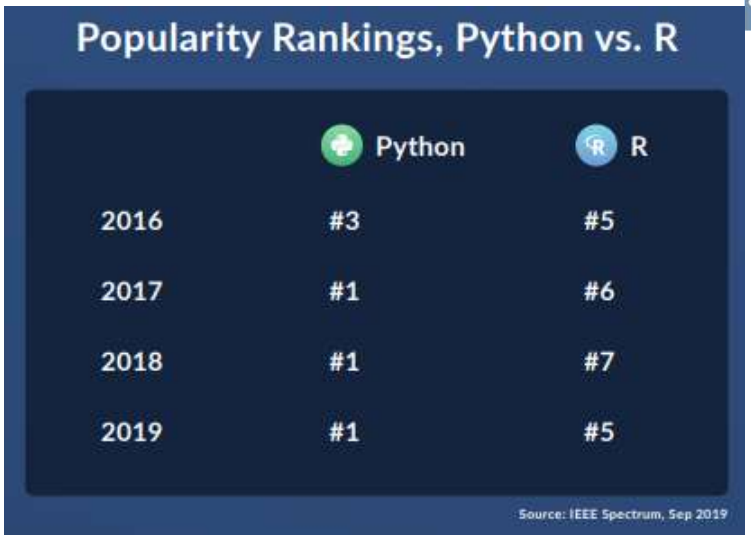
# Машинное обучение
# ???

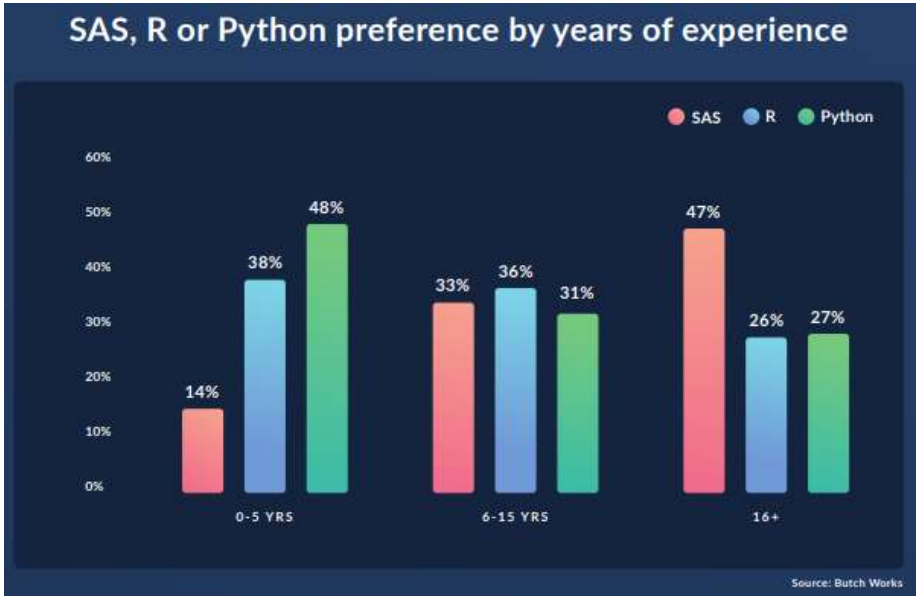


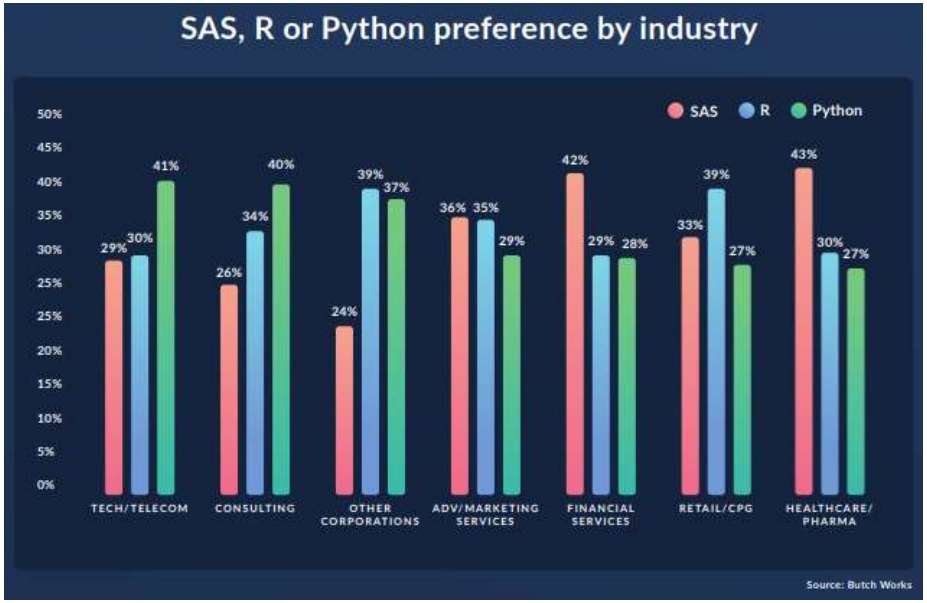

https://github.com/kozyrkov/presentations/blob/master/DecisionIntelligence.pdf

# R vs Python






https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis

# УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Дневная форма получения образования

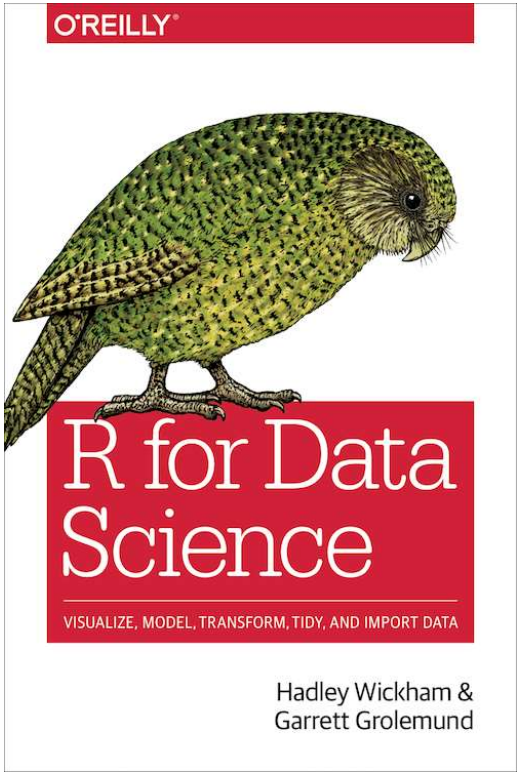
| Номер раздела, темы | Название раздела, темы | Количество аудиторных часов | | | | | Количество часов УСР | Форма контроля знаний |
|---|---|---|---|---|---|---|---|---|
| | | Лекции | Практические занятия | Семинарские занятия | Лабораторные занятия | Иное | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | Графический анализ данных | 2 | 4 | | | | | практическая работа 1 |
| 2 | Трансформация данных | 2 | 4 | | | | | практическая работа 2 |
| 3 | Исследовательский анализ данных | | 4 | | | | | практическая работа 3 |
| 4 | Основы машинного обучения. Регрессионный анализ данных | 4 | 6 | | | | | дискуссия, практическая работа 4 |
| 5 | Подготовка данных для моделирования. Модели классификации данных | 4 | 6 | | | | | дискуссия, практическая работа 5 |
| 6 | Модели кластеризации данных | 2 | 6 | | | | | практическая работа 6 |
| 7 | Индивидуальные и групповые проекты по анализу данных на R | | 4 | | | | | индивидуальный проект |
| | Итого | 14 | 34 | | | | | |

# Источники



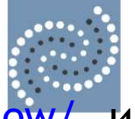
Темы: 1-3

«R for Data Science» https://r4ds.had.co.nz,
главы 1-8



# Тема 1. Графический анализ данных

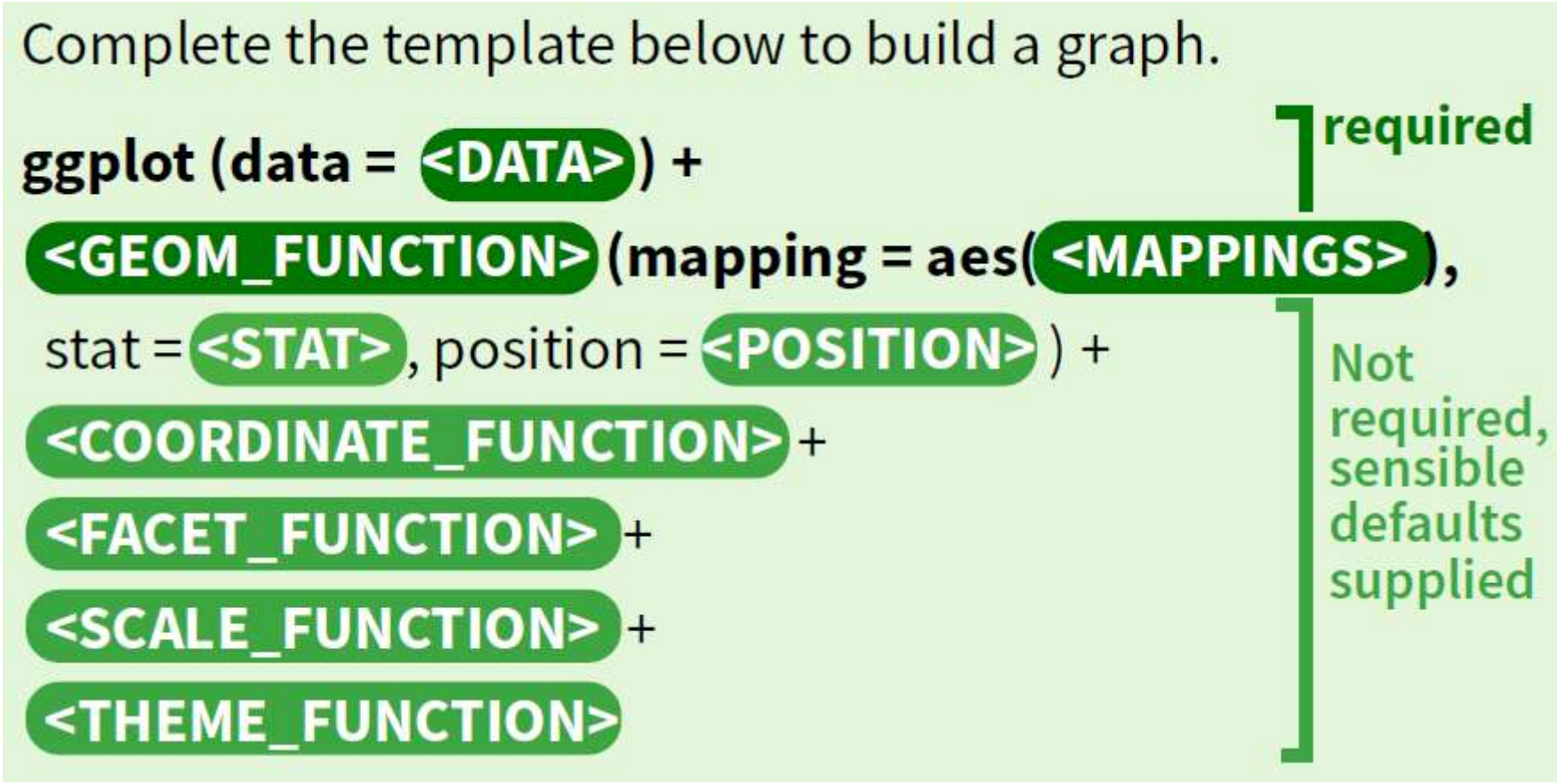




## Подготовка и начало работы

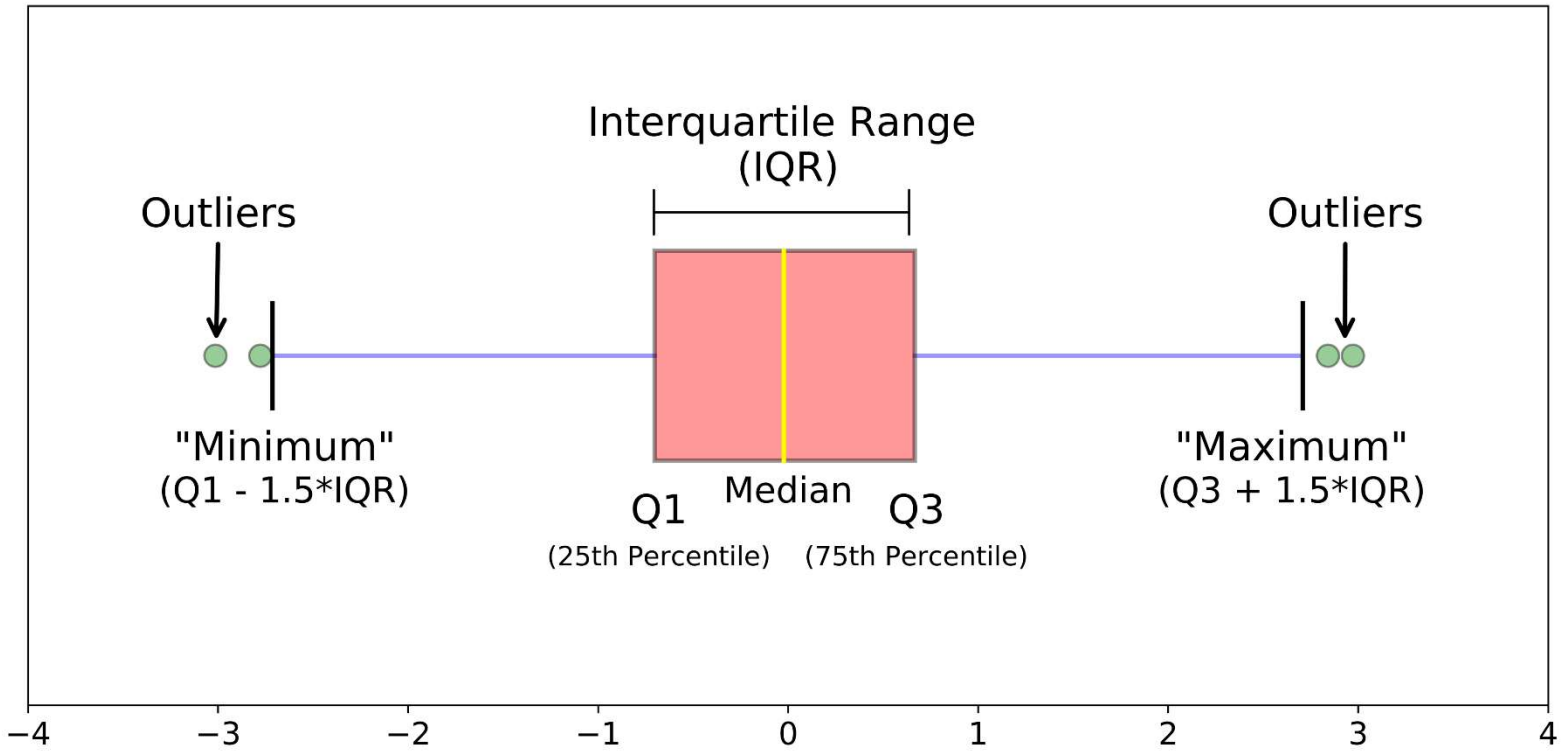
1. Ознакомиться с GitHub, прочитав статью https://guides.github.com/introduction/flow/, и создать на GitHub свой аккаунт (логин/пароль сохранить, например, в Evernote).
2. Прислать преподавателю (Telegram, user Kate) данные своего аккаунта (имя пользователя / емайл).
3. Выполнить упражнение Hello World по ссылке https://guides.github.com/activities/hello-world/. Осознать и обсудить в малых группах полученные результаты.
4. Принять приглашение на доступ к репозиторию курса, пришедшее на емайл.
5. Войти в репозиторий курса https://github.com/k-miniukovich/DA_BSU и создать branch (имя – ваша фамилия латинскими буквами).
6. Создать аккаунт в rstudio.cloud – облачная версия IDE RStudio (логин/пароль сохранить).
7. Создать в rsudio.cloud проект на базе github репозитория https://github.com/k-miniukovich/DA_BSU
8. Открыть созданный проект, перейти по закладке git (вверху справа) и выбрать свой branch (!!! Работаем только со своим branch)
9. Разобрать *1_visualization/visualization.R*
10. Изучить теоретический материал: «R for Data Science» https://r4ds.had.co.nz главы 1-4
11. Выполнить *pr_visualization/visualization.R*
12. Сделать Commit и Push из Rstudio в Github (свой branch). *Не выполнять merge* своего branch и master.

# Пакет ggplot2




Cheat Sheet https://rstudio.com/resources/cheatsheets/
Data Visualization with ggplot2

# Диаграмма размаха(boxplot)



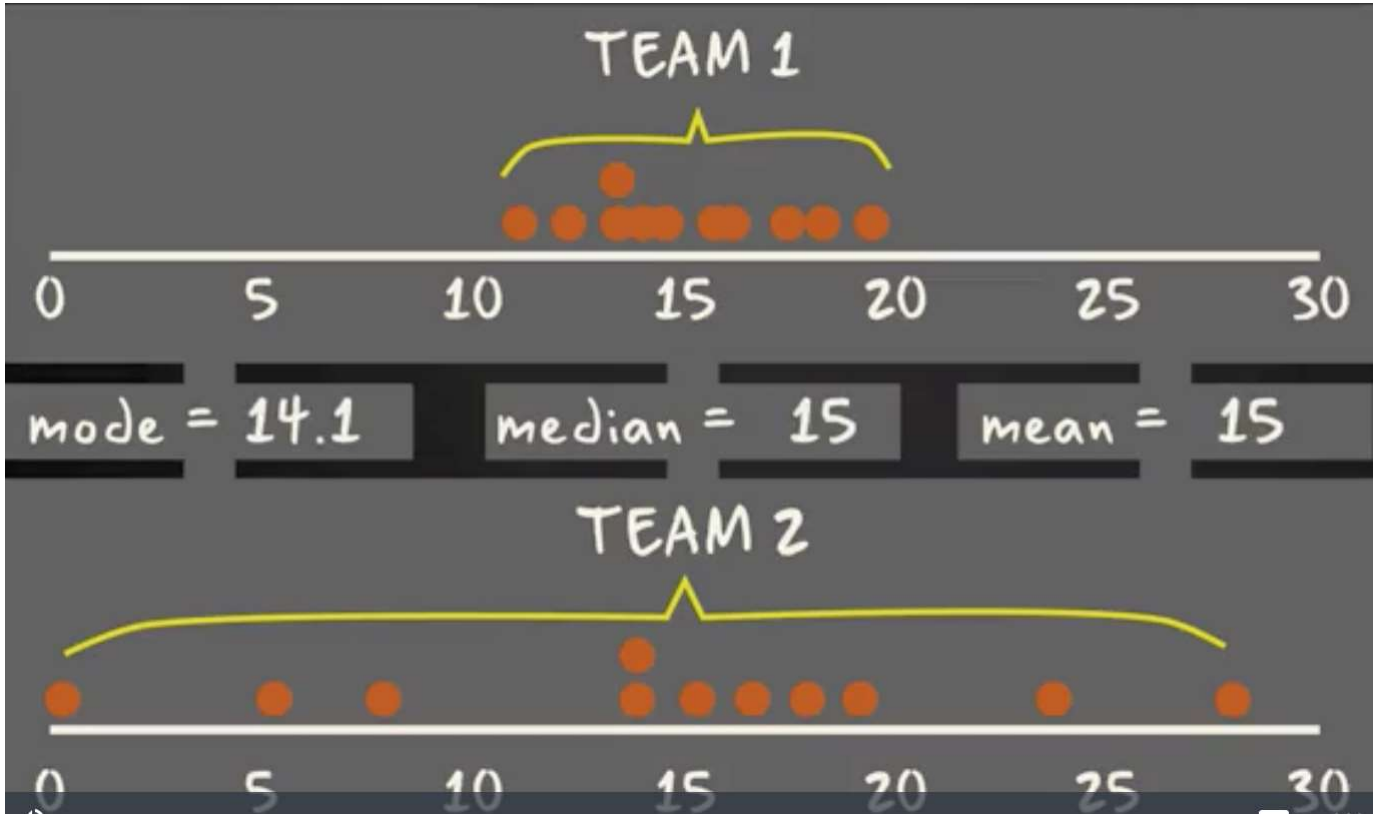

Ссылки

https://en.wikipedia.org/wiki/Box_plot#cite_note-4

https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51

https://en.wikipedia.org/wiki/Median

# Диаграмма размаха(boxplot)



visualization.R

- First steps
- Aesthetic mappings
- Facets
- Geometric objects
- Statistical transformations
- Sum up

Управляемая самостоятельная работа

pr_visualization.R

transformation.R

- filter()
- arrange()
- select()
- mutate()
- summarise()

Управляемая самостоятельная работа

pr_transformation.R

eda.R

- Variation
- Missing values
- Covariation
- Patterns

Управляемая самостоятельная работа

pr_eda.R