Разведочный анализ данных

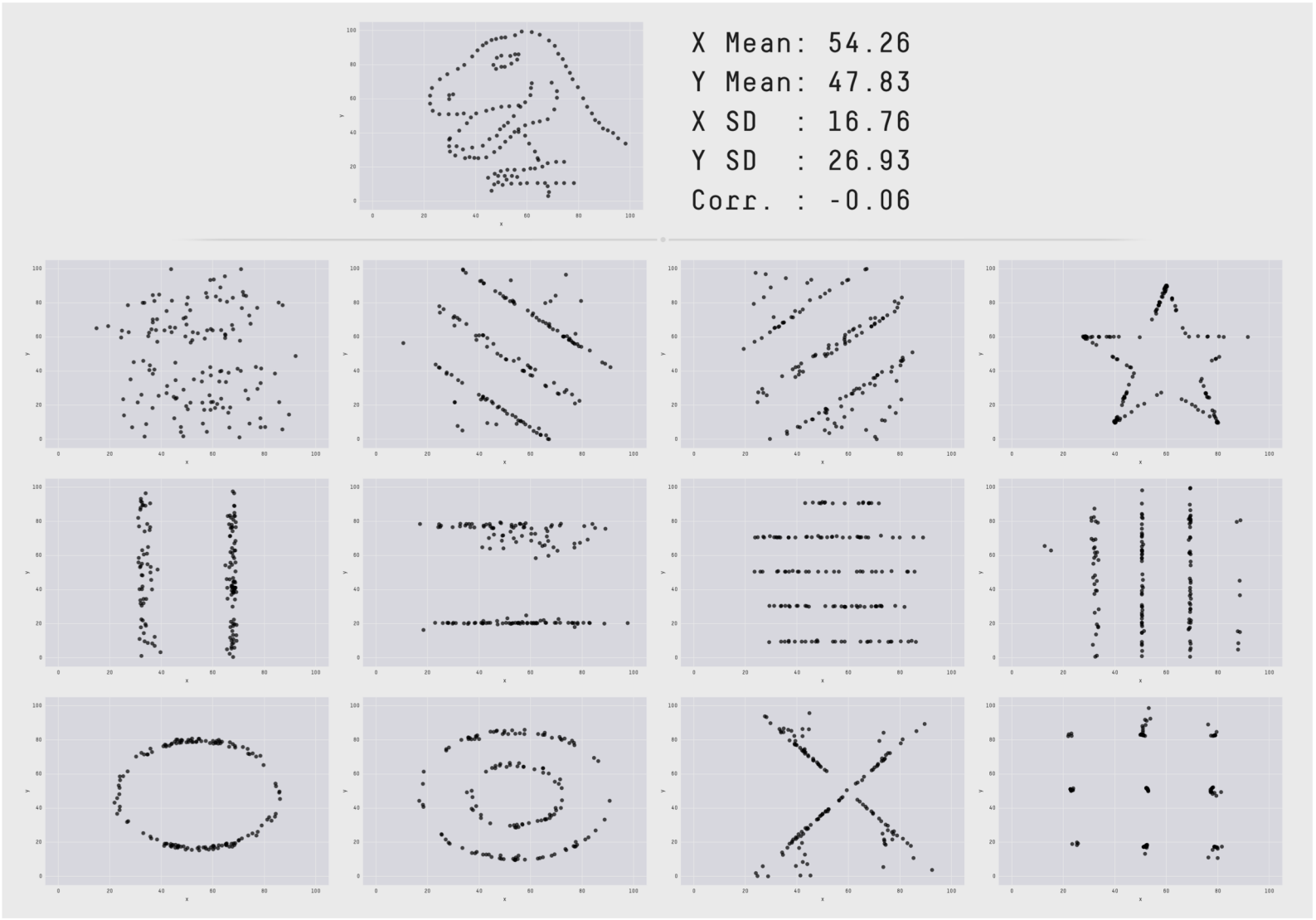
План

- Поговорим про разведочный анализ данных и их визуализацию
- Проанализируем данные о мемах
- Проанализируем данные о покупках в интернет-магазине

Зачем нужен разведочный анализ данных?

- Найти закономерности в данных, сформулировать гипотезы о новых закономерностях
- Помогает выявить основные проблемы в данных и понять как их лучше предобработать для обучения моделей
- Понимание природы данных поможет придумать новые признаки для обучения моделей

Зачем визуализировать данные?



➤ *https://www.autodesk.com/research/publications/same-stats-different-graphs*

Главные советы при подборе визуализаций

- Не используйте очень сложных средств визуализации (вы должны понимать что изображено на картинке)
- Разберитесь в достоинствах и недостатках каждого типа визуализации
- Не ограничивайтесь одной визуализацией
- Часто вместо всей выборки имеет смысл визуализировать случайную подвыборку (быстрее строится визуализация, лучше видна плотность скопления точек)