

What is statistics?

Jeff Leek

@jtleek

www.jtleek.com

What is statistics

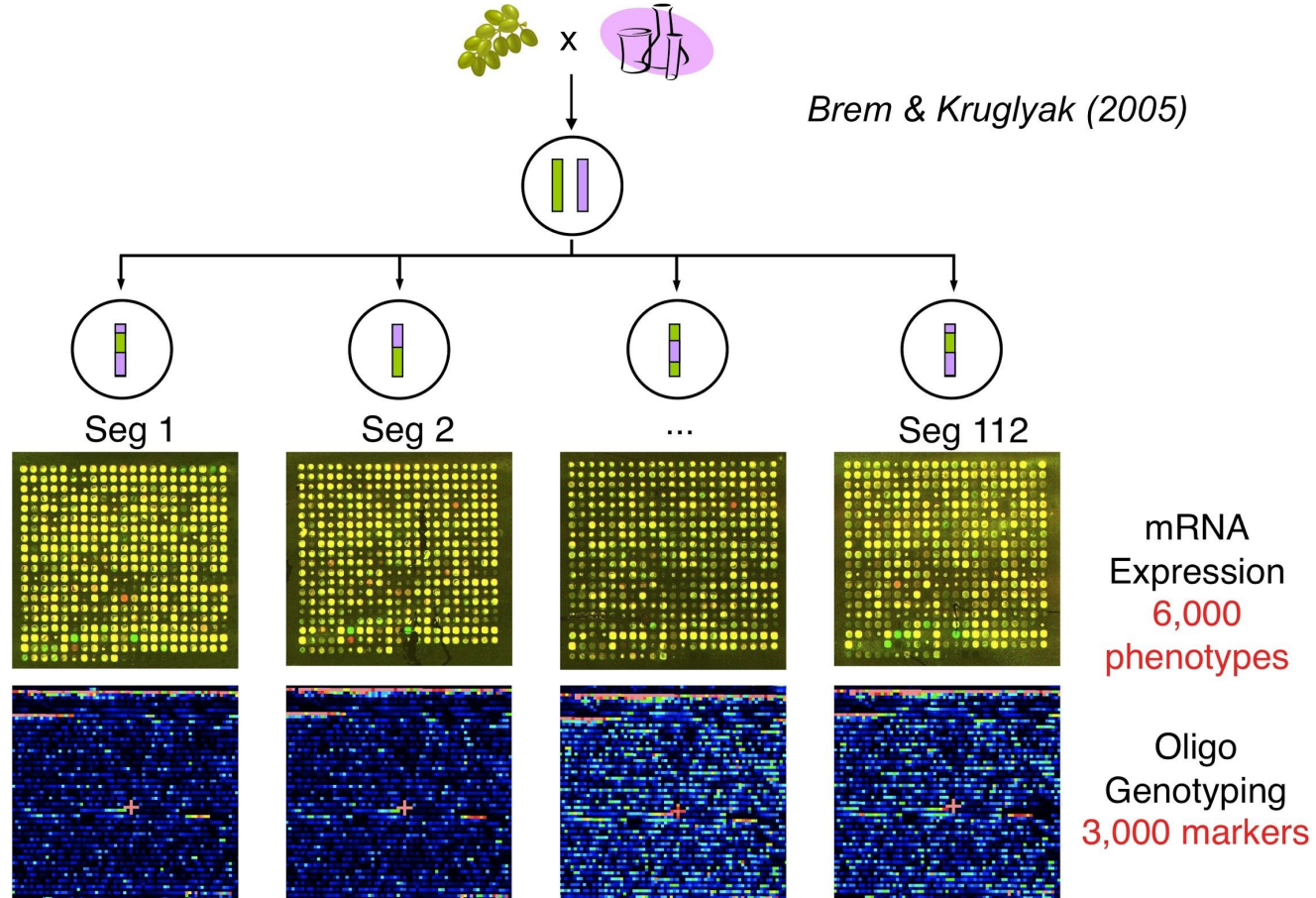
“

Statistics is the science of learning generalizable knowledge from data

”

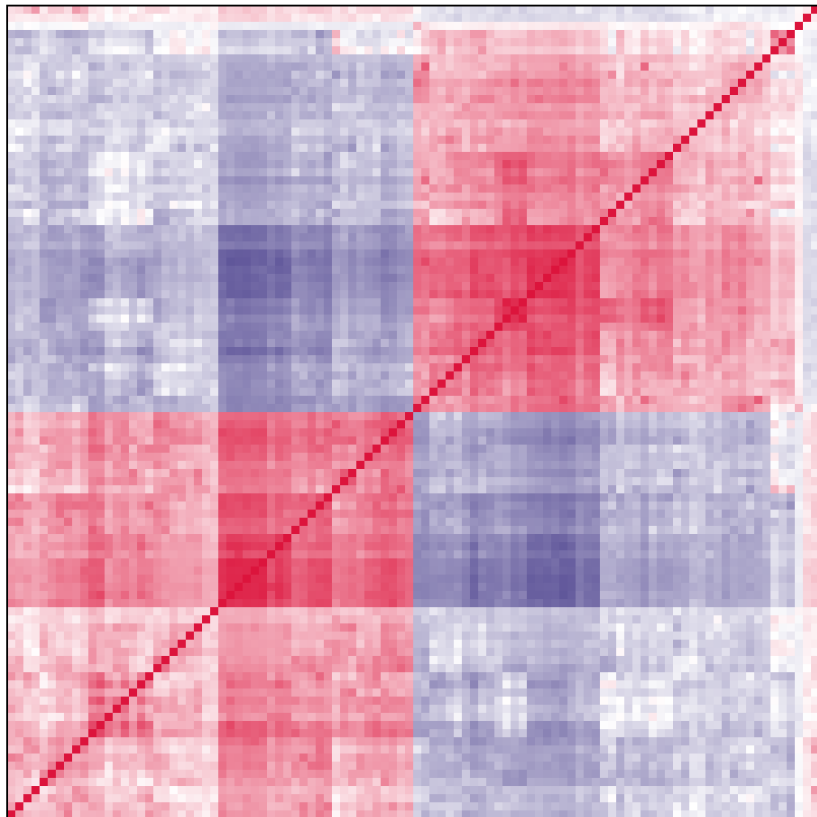
Statistics is study design

Genes and Genetic Variants

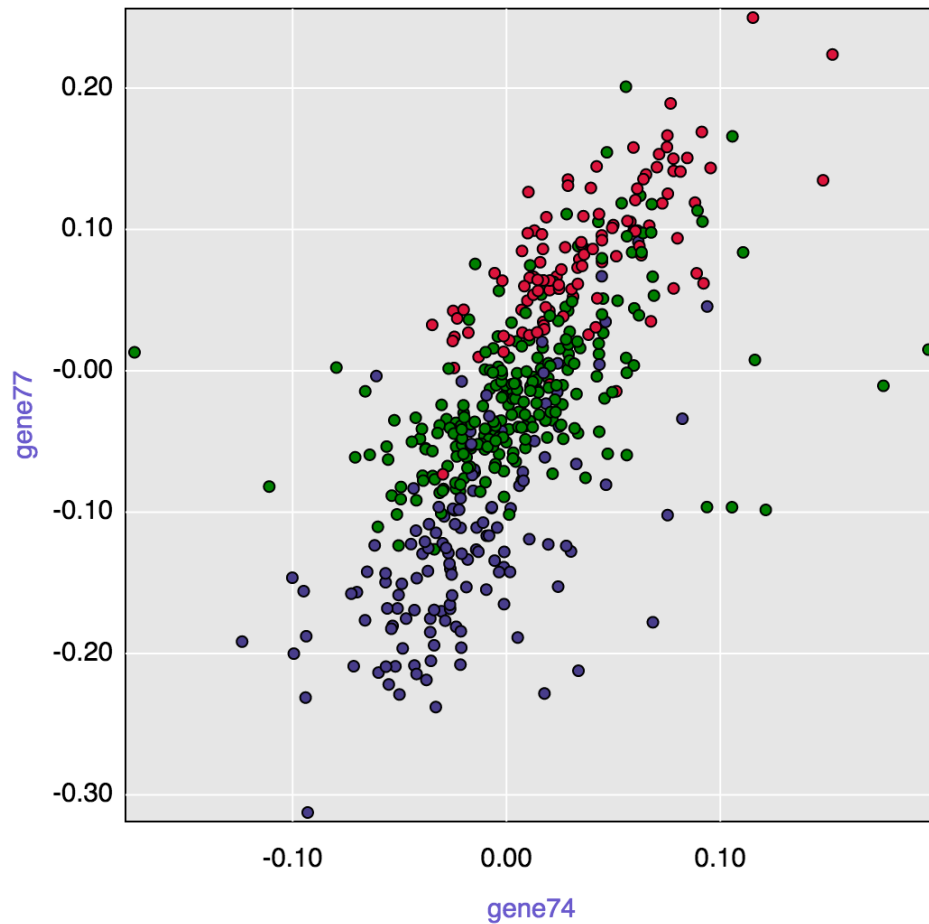


Statistics is exploratory analysis

Correlation matrix



Scatterplot



Statistics is pre-processing

Original

2	4	4
5	4	14
4	6	8
3	5	8
3	3	9



Ordered

2	3	4
3	4	8
3	4	8
4	5	9
5	6	14



Averaged

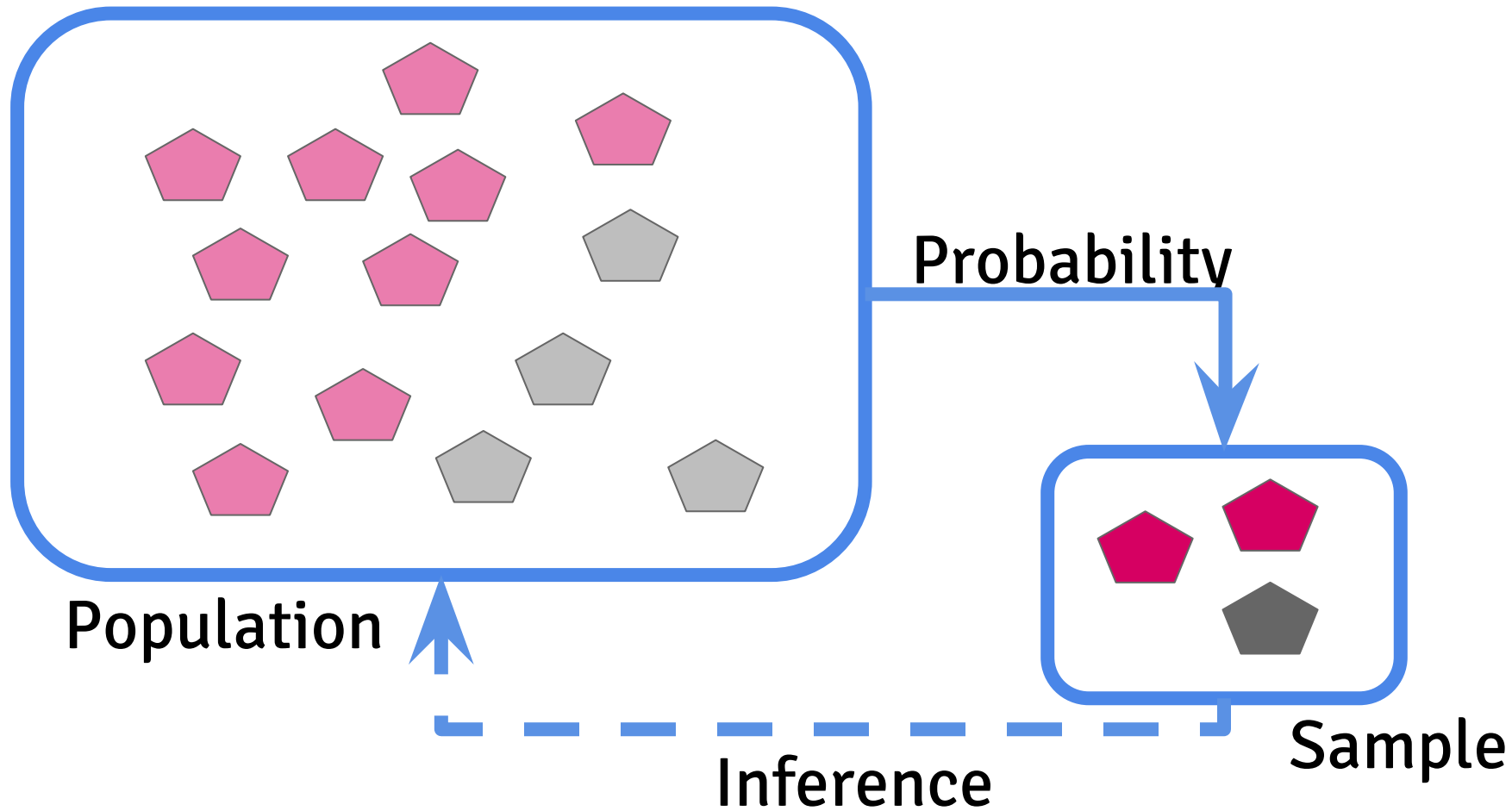
3	3	3
5	5	5
5	5	5
6	6	6
8	8	8



Re-ordered

3	5	3
8	5	8
6	8	5
5	6	5
5	3	6

Statistics is inference



Statistics is communication

Analysis of Recount data

<http://jtleek.com/svaseq/recount.html>

Jeff Leek

Load packages

You will need the RSkittleBrewer package for this vignette to run. Installation instructions are available here:

- <https://github.com/alyssafrazee/RskittleBrewer>

This code also accesses data from the ReCount database:

- <http://bowtie-bio.sourceforge.net/recount/>

You will also need R version 3.1.0 or greater and Bioconductor 3.0 or greater. The zebrafishRNASeq package might need to be installed from source. These analyses are based on the devel version of sva (version 3.11.2 or greater).

```
library(zebrafishRNASeq)
library(RSkittleBrewer)
library(genefilter)
library(RUVSeq)
library(edgeR)
library(sva)
library(ffpe)
```

```
## Warning: replacing previous import by 'graphics::image' when loading
## 'methyllumi'
```

```
library(RColorBrewer)
library(corrplot)
library(limma)
trop = RSkittleBrewer('tropical')
```