

Day 2 - outline

- Library/packages
- R basic plotting
- Exploratory data analysis and descriptive statistics

Libraries / packages

Library/packages

- **Packages** are collections of R functions, data, and compiled code in a well-defined format.
- The directory where packages are stored is called the **library**.

Definitions from <http://www.statmethods.net/interface/packages.html>

R-base: standard packages

- **R-base:**

about 25 standard packages supplied with R by default (example: base, stats, graphics).

R-contrib: all other packages

- 2 main repositories:
 - CRAN: Comprehensive R Archive Network
11540 packages available
<https://cran.r-project.org>
 - Bioconductor:
1383 packages available
<https://www.bioconductor.org/>

Find packages

- R packages in CRAN:

<https://cran.r-project.org/web/packages/>

Available CRAN Packages By Date of Publication

Date	Package	Title
2017-05-21	aSPU	Adaptive Sum of Powered Score Test
2017-05-21	d3r	'd3.js' Utilities for R
2017-05-21	dynr	Dynamic Modeling in R
2017-05-21	cdfReader	Reading EDF(+) and BDF(+) Files
2017-05-21	GlobalOptions	Generate Functions to Get or Set Global Options
2017-05-21	GWmodel	Geographically-Weighted Models
2017-05-21	IgorR	Read Binary Files Saved by 'Igor Pro' (Including 'Neuromatic' Data)
2017-05-21	imputeTS	Time Series Missing Value Imputation
2017-05-21	lagged	Classes and Methods for Lagged Objects
2017-05-21	mclust	Gaussian Mixture Modelling for Model-Based Clustering, Classification, and Density Estimation
2017-05-21	NPMOD	Non Parametric Module
2017-05-21	numGen	Number Series Generator
2017-05-21	penRvine	Flexible R-Vines Estimation Using Bivariate Penalized Splines
2017-05-21	PhenotypeSimulator	Flexible Phenotype Simulation from Different Genetic and Noise Models
2017-05-21	plfMA	A GUI to View, Design and Export Various Graphs of Data
2017-05-21	RANN	Fast Nearest Neighbour Search (Wraps ANN Library) Using L2 Metric
2017-05-21	regnet	Network-Based Regularization for Generalized Linear Models
2017-05-21	timereg	Flexible Regression Models for Survival Data
2017-05-20	AIG	Automatic Item Generator

Find packages

- R packages in Bioconductor:
<https://bioconductor.org/packages>

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Home Install Help Developers About

Search:

Home » BiocViews

All Packages

Bioconductor version 3.5 (Release)

Autocomplete biocViews search:

▼ Software (1381)

- ▶ AssayDomain (525)
- ▶ BiologicalQuestion (507)
- ▶ Infrastructure (297)
- ▶ ResearchField (373)
- ▶ StatisticalMethod (441)
- ▶ Technology (872)
- ▶ WorkflowStep (735)
- ▶ AnnotationData (912)
- ▶ ExperimentData (316)

Packages found under Software:

Show entries

Search table:

Package	Maintainer	Title
a4	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Umbrella Package
a4Base	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Base Package
a4Classif	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Classification Package
a4Core	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Core Package
a4Preproc	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Preprocessing Package
a4Reporting	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Reporting Package
ABAEnrichment	Steffi Grote	Gene expression enrichment in human brain regions
...	Yongming Andrew	Microarray QA and statistical data analysis for Applied Biosystems Genome

Bioconductor

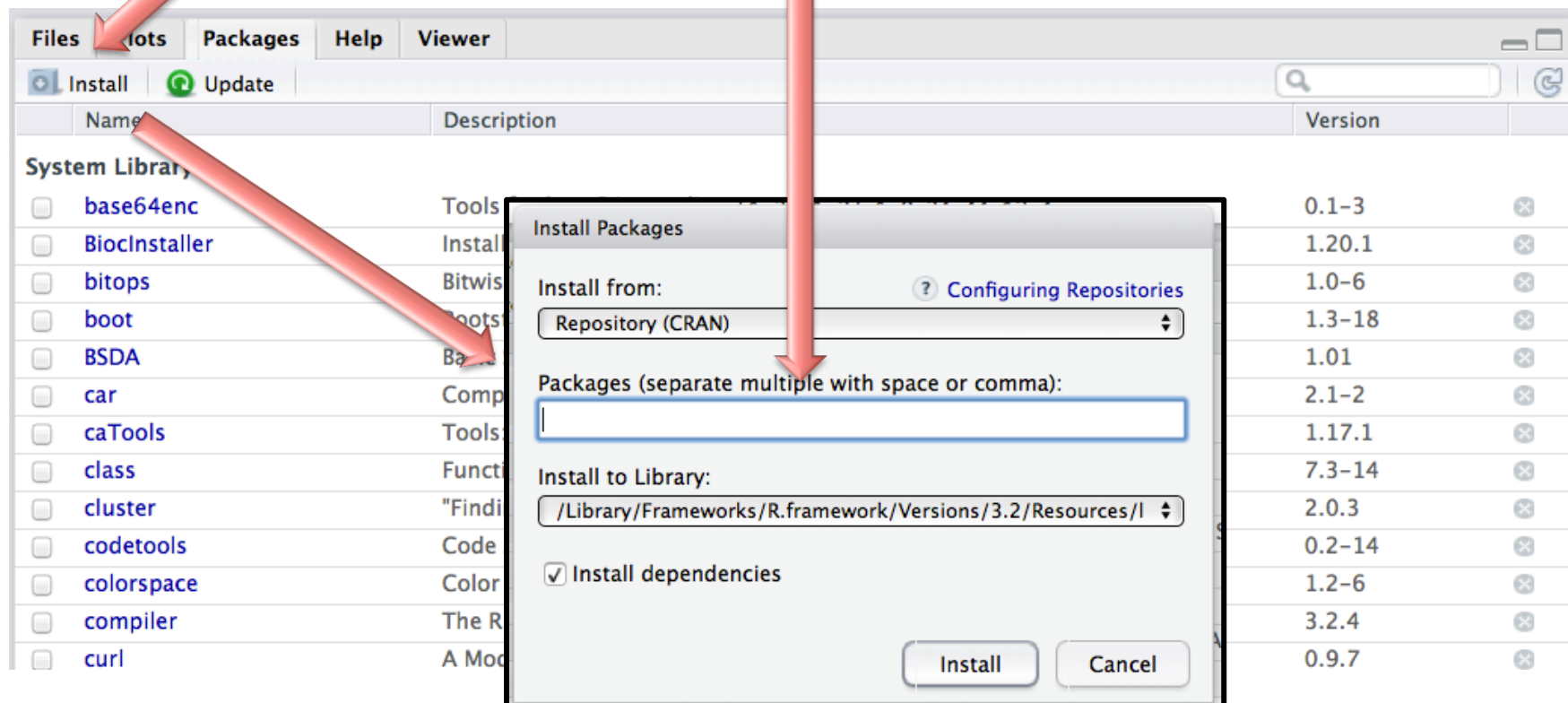
- R packages specialized in bioinformatics analysis
 - Supports most types of genomics and NGS data (e.g. limma, DESeq2, BayesPeak)
 - Specific data classes (e.g. Granges from GenomicRanges)
 - Integrates command line tools (e.g Rsamtools)
 - Annotation tools (e.g. biomaRt)

Types of Bioconductor packages

- **Software:** set of functions
e.g. DESeq2(NGS data analysis)
- **Annotation:** annotation of specific arrays, organisms, events, etc.
e.g. BSgenome.Hsapiens.UCSC.hg38
- **Experiment:** data that can be loaded and used
e.g. ALL (acute lymphoblastic leukemia dataset)

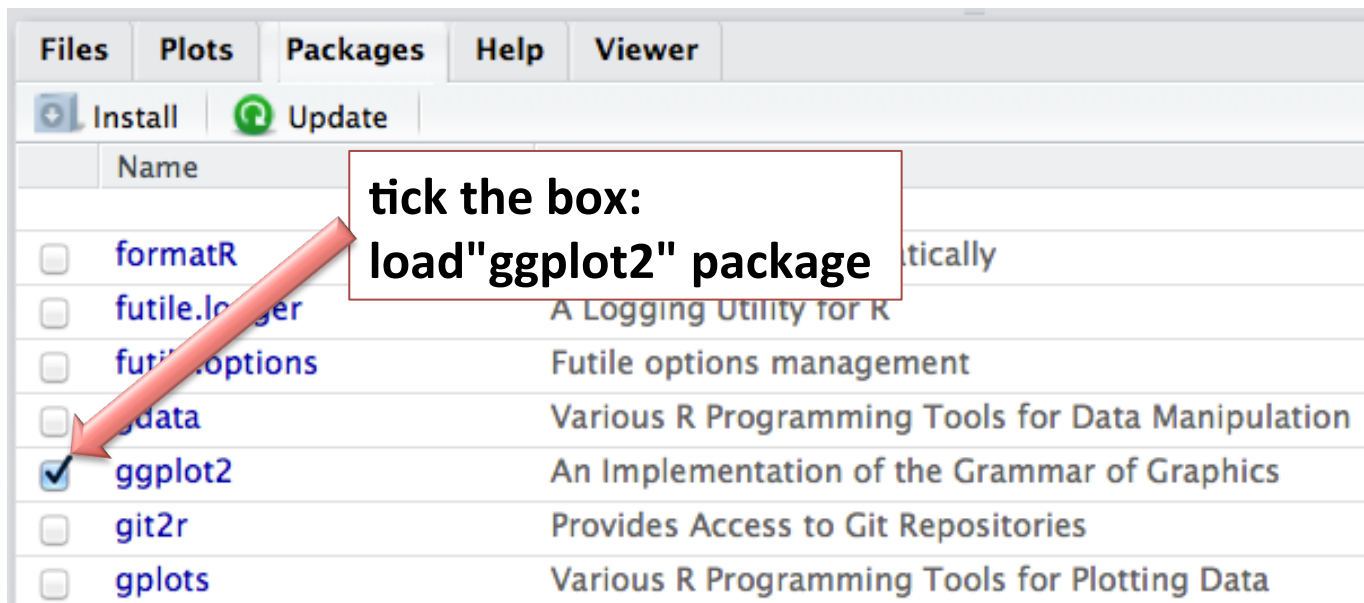
Install a package with R studio

Install "ggplot2" package



Load a package

- **from console**
`library(ggplot2)`
- **with RStudio**



Listing functions from packages

- `ls("package:ggplot2")`

```
[1] "%+%"           "+replace%"
[3] "Coord"         "CoordCartesian"
[5] "CoordFixed"    "CoordFlip"
[7] "CoordMap"      "CoordPolar"
[9] "CoordQuickmap" "CoordTrans"
[11] "Geom"          "GeomAblin"
[13] "GeomAnnotationMap" "GeomArea"
[15] "GeomBar"       "GeomBlank"
```

...

Graphing in R:

basic graphing

Producing graphs

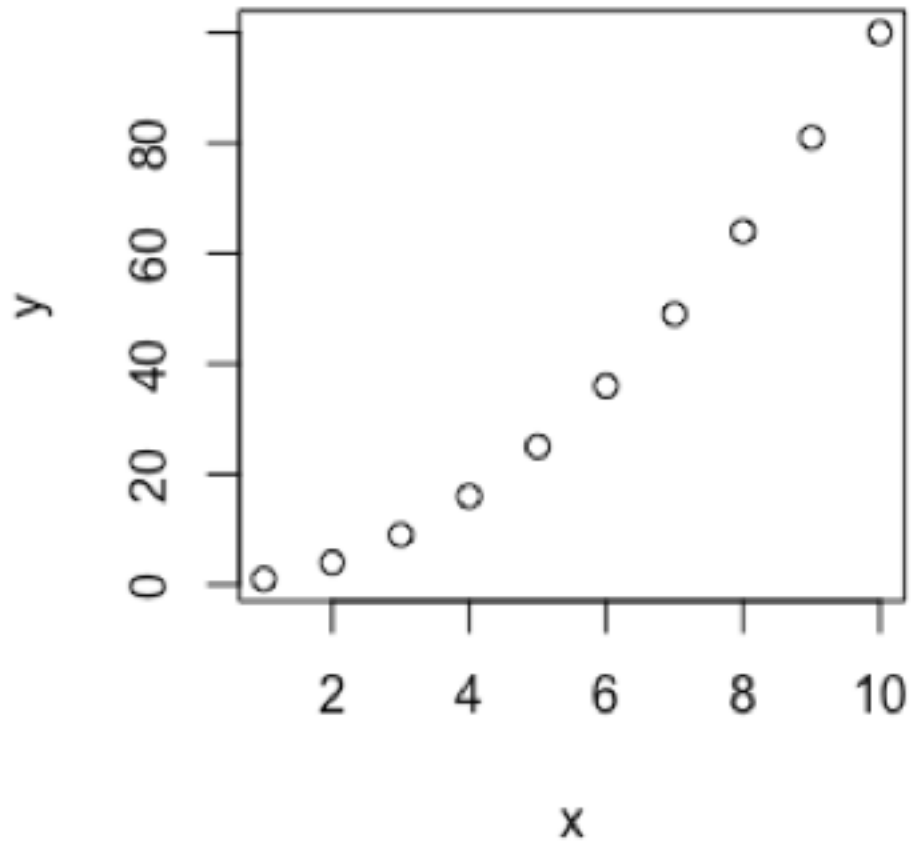
- R-base package **graphics** offers functions for producing many plots, for example:
 - scatter plots – `plot()`
 - histograms – `hist()`
 - boxplots – `boxplot()`
 - bar plots – `barplot()`
 - pie charts – `pie()`

Scatter plots

A scatter plot has points that show the **relationship** between two sets of data.

Basic scatter plot

- $x \leftarrow 1:10$
- $y \leftarrow x^2$
- **plot(x, y)**



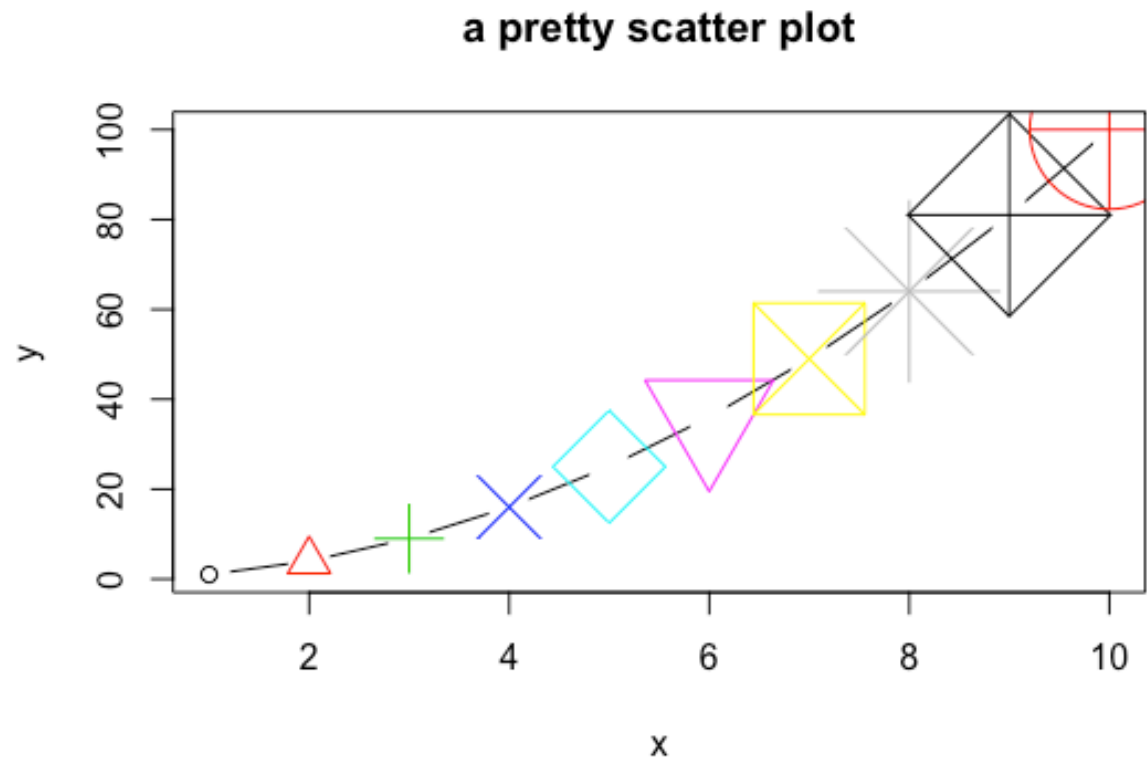
Scatter plot

- `plot(x, y, col="red", pch=2, type="b", main="a pretty scatter plot")`



Scatter plot

- `plot(x, y, col=1:10, pch=1:10, cex=1:10, type="b", main="a pretty scatter plot")`

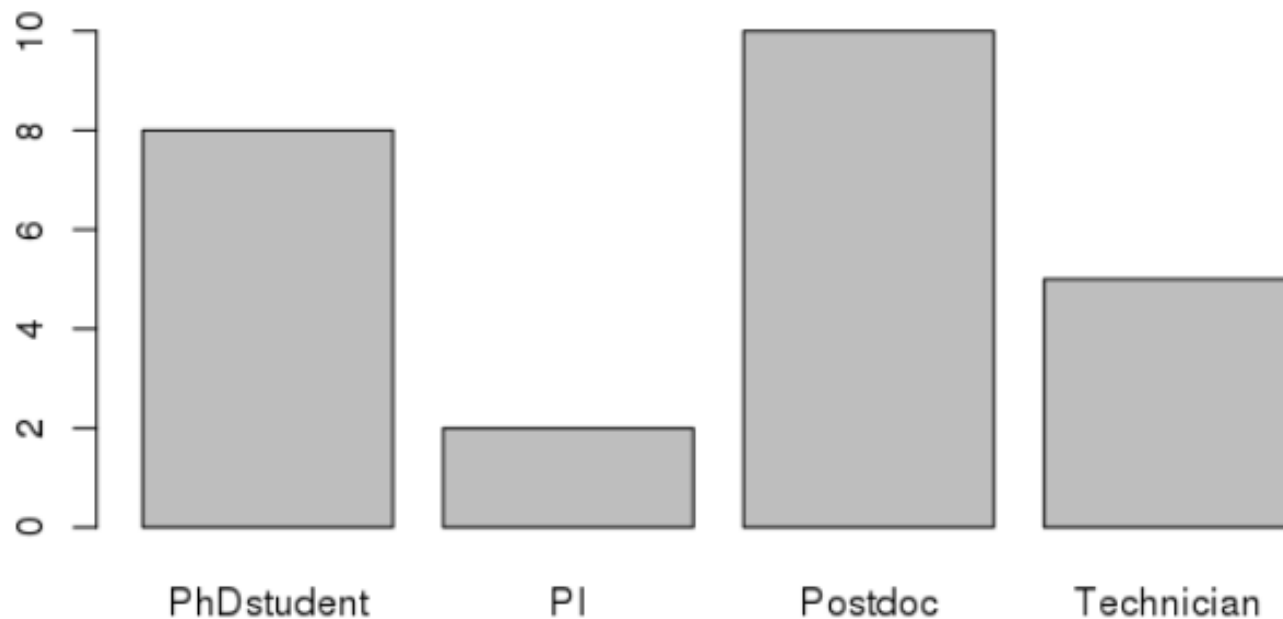


Barplots

A bar chart or bar plot displays rectangular bars with lengths **proportional** to the values that they represent.

Barplot

- `x <- rep(c("PhDstudent", "Postdoc", "Technician", "PI"), c(8,10,5,2))`
- `barplot(table(x))`



Barplot

- Change order of blocks and y axis ticks

reorder vector with its names

```
y <- table(x)[c("PI", "Technician", "PhDstudent", "Postdoc")]
```

remove axes

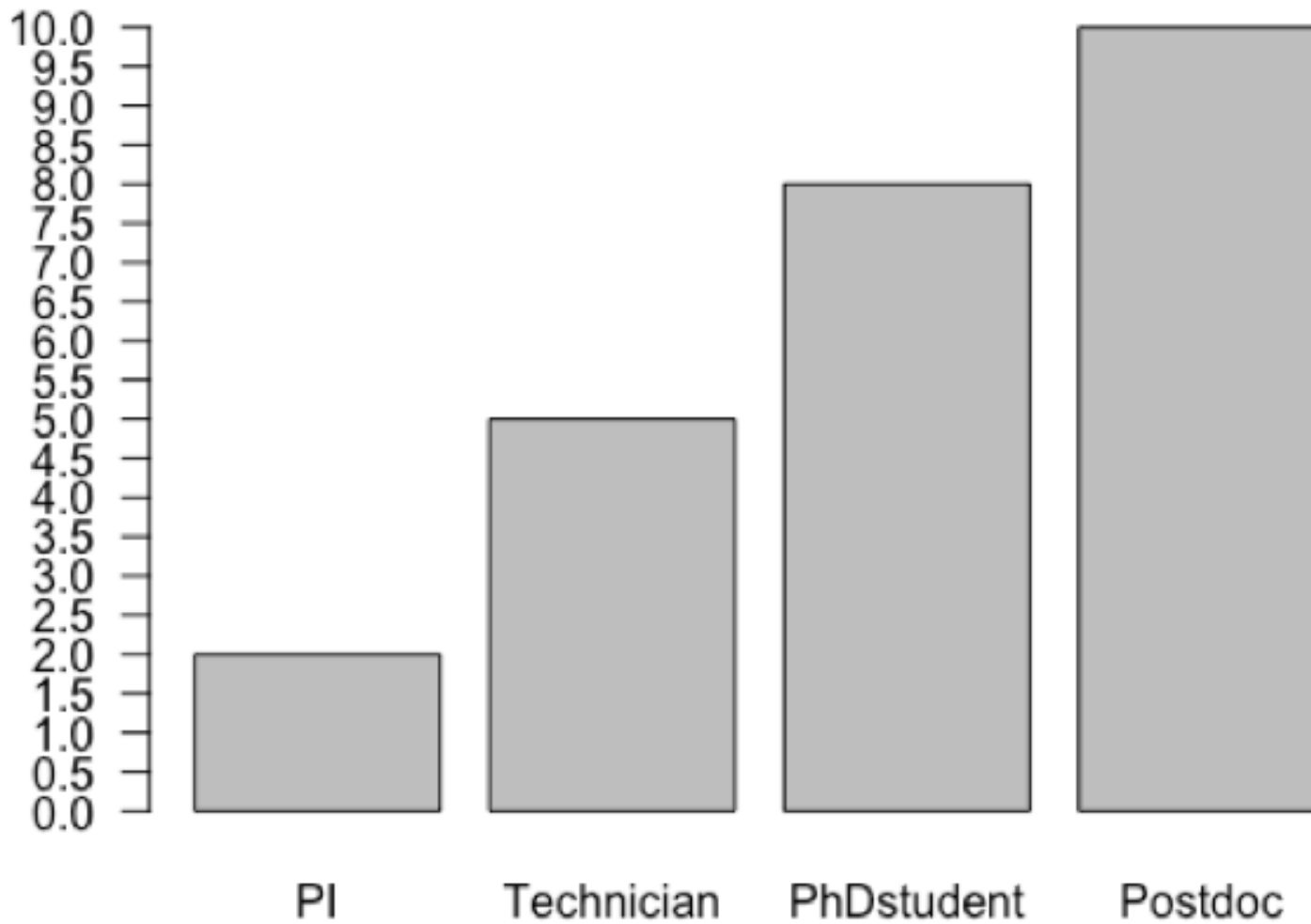
- **barplot**(y, axes=F)

add new y axis with specific ticks

las=2 displays tick labels horizontally

- `axis(2, at = seq(0,10, 0.5), las=2)`

Barplot

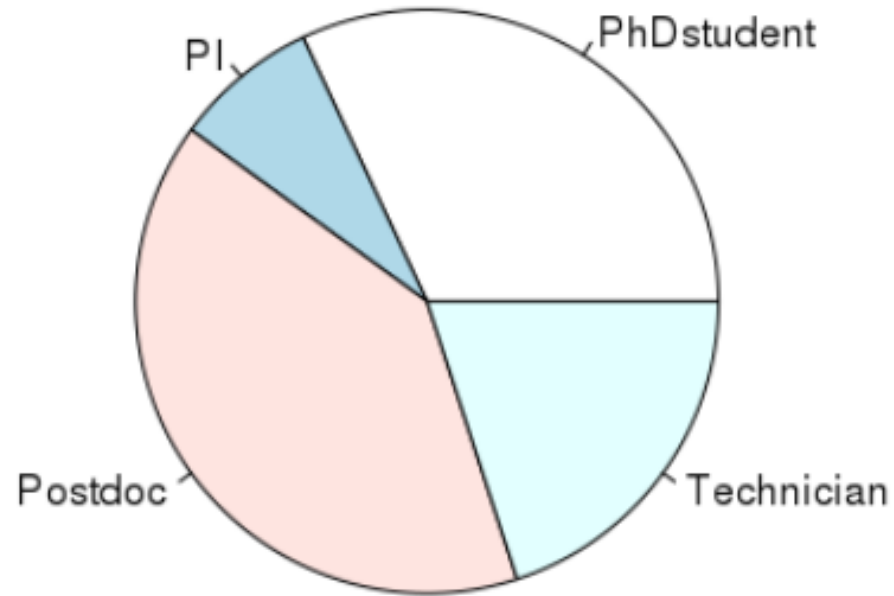


Pie chart

A pie chart is a circular charts which is divided into slices, illustrating **proportions**.

Pie chart

- `x <- rep(c("PhDstudent", "Postdoc", "Technician", "PI"), c(8,10,5,2))`
- **pie(table(x))**

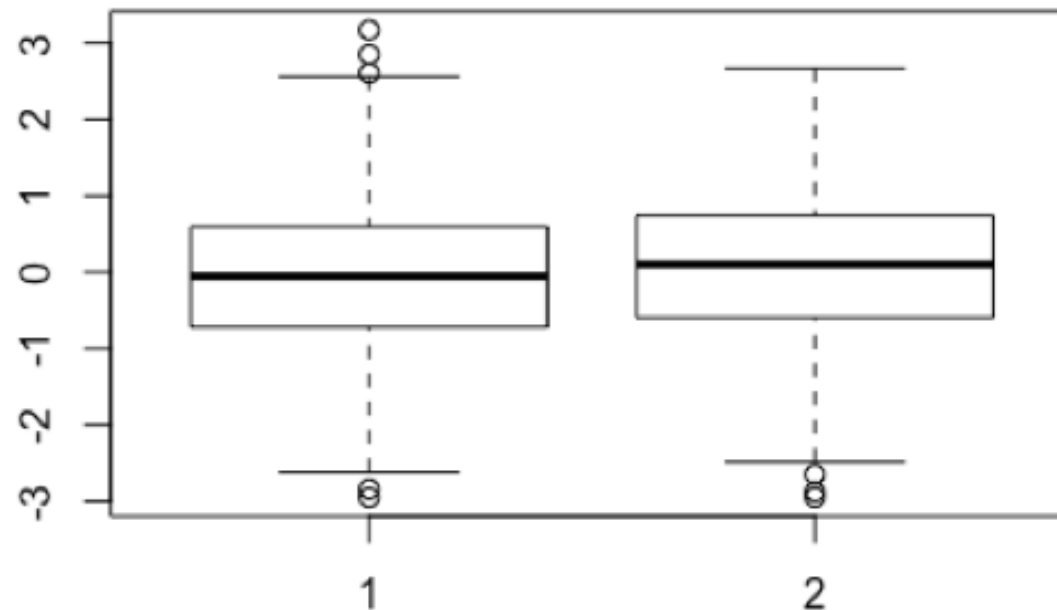


Boxplots

A boxplot is a convenient way to describe the **distribution** of the data.

Boxplots

- `x <- matrix(rnorm(1000), ncol=2)`
- `boxplot(x)`

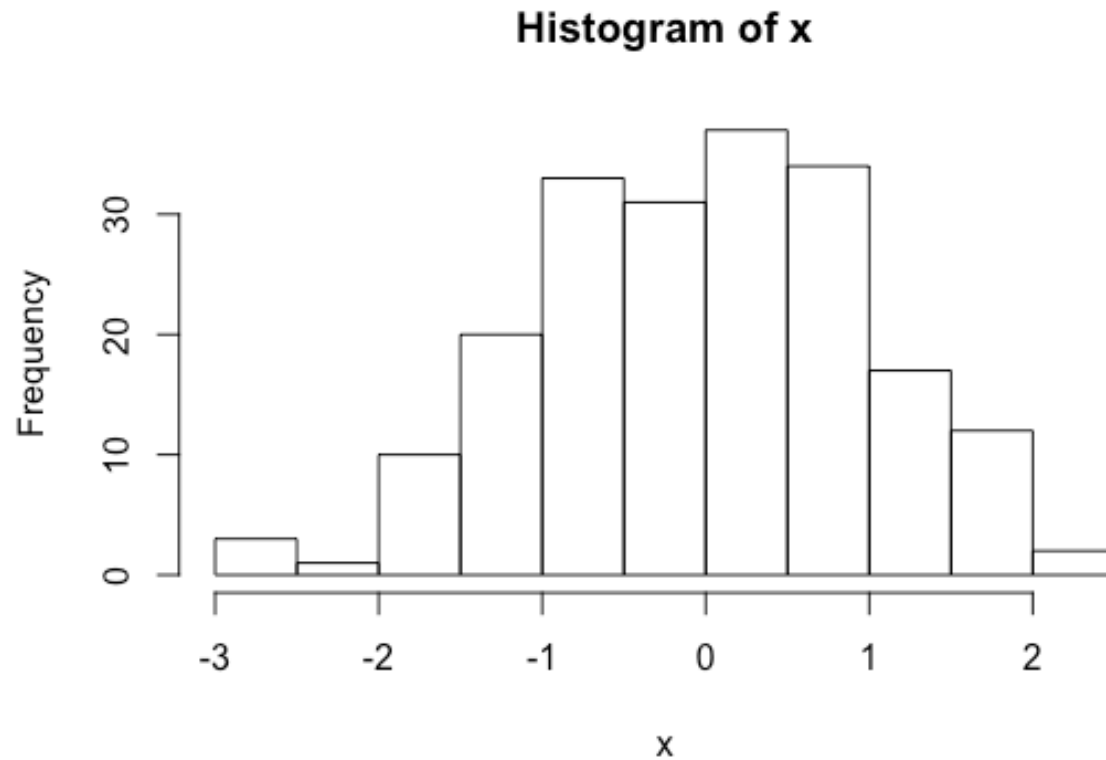


Histograms

A histogram graphically summarizes the **distribution** of the data.

Histogram

- `x <- rnorm(200)`
- **`hist(x)`**



Exporting graphs

- Graphs can be saved into many formats, including: **pdf, jpeg, bmp, tiff**.
- Open the file to save your plot in:
`pdf("my_graph.pdf")`
- Produce a graph:
`plot(1:10)`
- And close it:
`dev.off()`

Exporting graphs

- From R Studio:

