R basics
Assign number 5 to variable x

\[
\text{x} \leftarrow 5
\]

\[
\text{x}
\]

\[
[1] 5
\]

Calculate \(5 \times x^2 + 7\)

\[
5 \times x^2 + 7
\]

\[
[1] 132
\]

Create vector, assign to variable y

\[
\text{y} \leftarrow \text{c}(1, 2, 3, 4, 5)
\]

\[
y
\]

\[
[1] 1 2 3 4 5
\]

Multiply each element in y with the number in x

\[
x \times y
\]

\[
[1] 5 10 15 20 25
\]
A string contains text:

> name <- "Claus Wilke"
> name
[1] "Claus Wilke"

A vector of strings:

> animals <- c("cat", "mouse", "mouse", "cat", "rabbit")
> animals
[1] "cat"    "mouse"  "mouse"  "cat"    "rabbit"
Factors

Factors keep track of distinct categories (levels) in a vector:

> animals
[1] "cat"   "mouse" "mouse" "cat"   "rabbit"

> factor(animals)
[1] cat    mouse  mouse  cat    rabbit
Levels: cat mouse rabbit
Data frames

We use data frames to store data sets with multiple variables:

```r
> pets <- data.frame(
    family = c(1, 2, 3, 4, 5),
    pet = animals
 )

> pets
   family pet
1     1  cat
2     2 mouse
3     3 mouse
4     4  cat
```
We access individual columns in a data frame with $ + the column name:

```r
> pets$family
[1] 1 2 3 4 5
```

```r
> pets$pet
[1] cat  mouse mouse cat  cat  rabbit
Levels: cat mouse rabbit
```
R has many built-in data frames:

```r
> cars
    speed dist
   1    4    2
   2    4   10
   3    7    4
   4    7   22
   5    8   16
   6    9   10
   7   10   18
   8   10   26
   9   10   34
  10   11   17
```
The `head()` function shows the first few lines of a data frame:

```r
> head(cars)
speed  dist
1     4   2
2     4   10
3     7    4
4     7   22
5     8   16
6     9   10
>
Hypothesis testing: a quick review
$H_0$ and $H_A$: Null and alternative hypothesis

$H_0$: Null hypothesis, assumption that the data show no signal, that nothing has happened.

$H_A$: Alternative hypothesis, opposite of $H_0$, assumption that something has happened.
The $P$ value tells us how unexpected the data are

**P value**: Probability to observe the given data under the assumption that $H_0$ is true

We generally reject $H_0$ if $P < 0.05$

We *never* accept $H_A$
**t test**: Do two groups of numerical measurements have the same mean?
Correlation: Do two numerical variables have a relationship with each other?
Multivariate regression: Which predictors have an effect on the response variable?

Example: