Workshop: Working with Strings in R

Simon Ress | Ruhr-Universität Bochum

October 18, 2021

Content

1. Base R
2. Package: Stringr
3. Helpful sources

Base R

Overview

The primary R functions for dealing with strings:

- substr()/substring(): Extract or replace substrings in a character vector
- strsplit(): Split the elements of a character vector x into substrings according to a given character
- paste(): concatenate n number of strings
- nchar(): returns a vector of the number of characters of x
- Dealing wirth regular expressions:
  - grep(), grepl(): These functions search for matches of a regular expression/pattern in a character vector.
  - regexpr(), gregexpr(): Search a character vector for regular expression matches and return the indices of the string where the match begins and the length of the match
  - sub(), gsub(): Search a character vector for regular expression matches and replace that match with another string

Pattern matching and replacement

grep(): Index in vector which matches regex

```
grep("b+", c("abc", "bda", "cca a", "abd"))
```

```
## [1] 1 2 4
```

regexpr(): Search a character vector for regular expression matches and return the indices of the string where the match begins and the length of the match

```
str = "Line 129: O that this too too solid flesh would melt
regexpr("1",str)
```

```
## [1] 6
## attr(,"match.length")
## [1] 1
## attr(,"index.type")
```

Matching

substr()/substring(): Extract or replace substrings in a character vector

```
num <- "12345678"
substr(num, 4, 5)
```

```
## [1] "45"
```

```
substring(num, 1:3, 7)
```

```
## [1] "1234567" "234567"  "34567"
```

Other

strsplit(): Split the elements of a character vector x into substrings according to a given character

```
str = "Splitting sentence into words"
strsplit(str, " ")
```

```
## [[1]]
## [1] "Splitting" "sentence"  "into"      "words"
```

paste(): concatenate n number of strings

```
paste("Count number", "of characters")
```

```
## [1] "Count number of characters"
```

nchar(): returns a vector of the number of characters of x

Package: Stringr

String length

Matching patterns with regular expressions

Process Matches

- Bullet 1
- Bullet 2
- Bullet 3

Helpful sources

Helpful sources

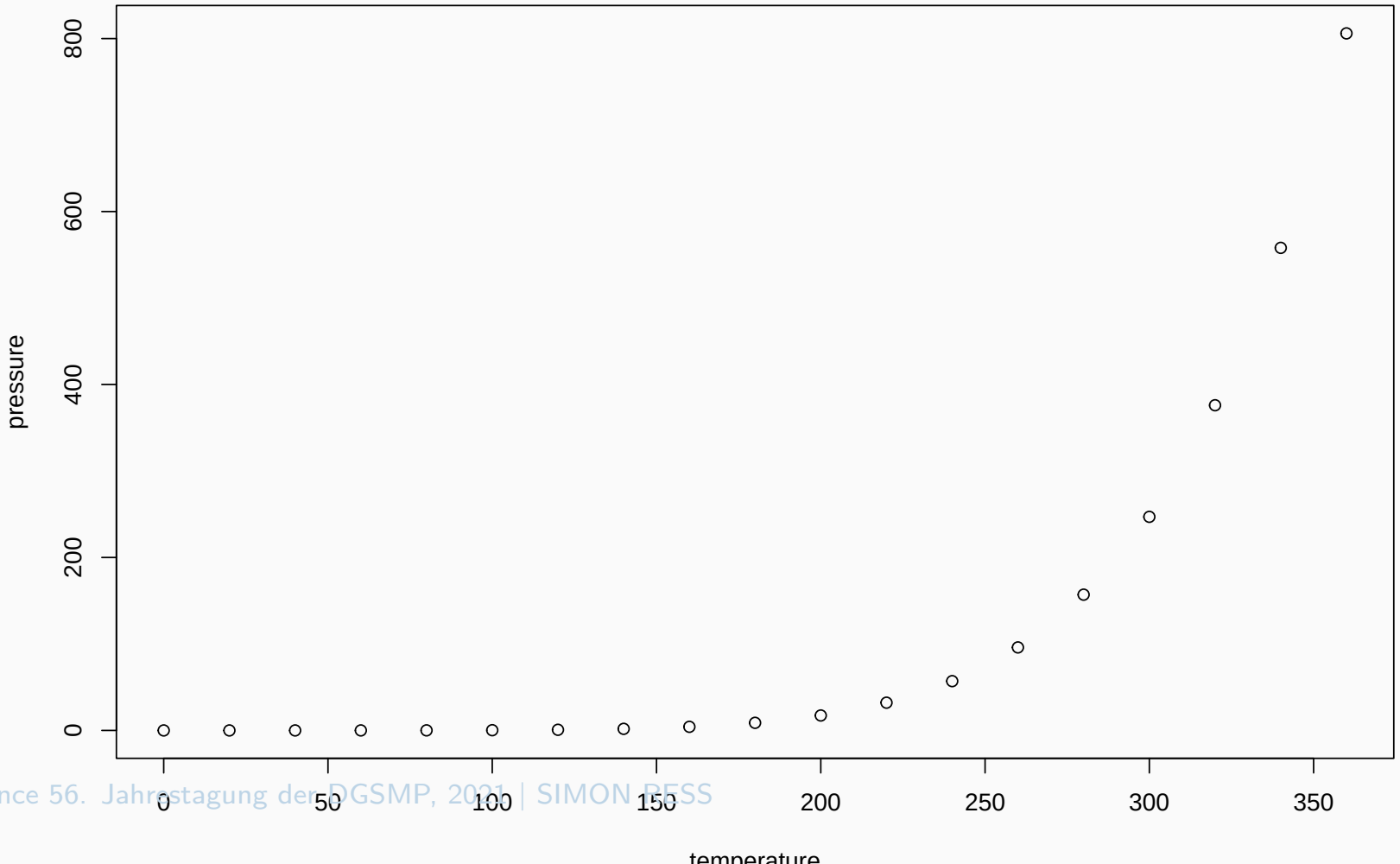
- Stringr: Introduction
- Stringr: Cheatsheet
- Stringr: Reference manual
- Base R String-functions vs Stringr
- Regular expressions
- Primary R functions for dealing with regular expressions

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Slide with Plot

```
plot(pressure)
```

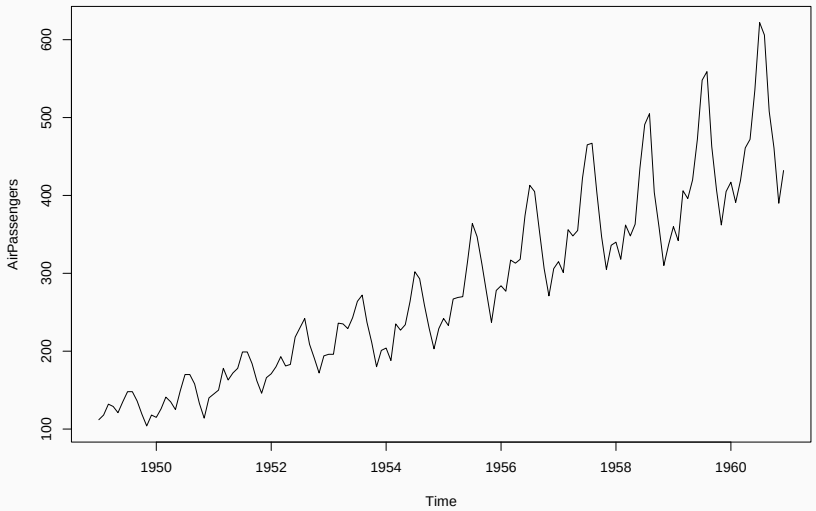


Two column layout

Here is some text above which goes over to whole slide

```
plot(AirPassengers)
```

- Description of plot
- Second point



and here some text below which goes over to whole slide

Breakout page

Using LaTeX Parts: Blocks

As one example of falling back into LaTeX, consider the example of three different block environments are pre-defined and may be styled with an optional background color.

Default

Block content.

Alert

Block content.

Example

Block content.