String Manipulation

with BaseR and TidyVerse(`stringr`)

String Functions

BaseR	stringr
tolower	str_c
toupper	str_length
substring	str_sub
paste	str_trim

```
library(tidyverse)

s0 <- "  The quick agile fox jumps over the very lazy dog "
s0
```

```
[1] "  The quick agile fox jumps over the very lazy dog "
```

```
str_length(s0)
```

```
[1] 51
```

```
s1 <- str_trim(s0, side="both")
s1
```

```
[1] "The quick agile fox jumps over the very lazy dog"
```

```
str_length(s1)
```

```
[1] 48
```

---

strsplit (returns a list)

```
# this happens with website crawling
t0 <- "get%better%every%day"
t1 <- strsplit(t0,split='%')
t1
```

```
[[1]]
[1] "get"    "better" "every"  "day"
```

```
typeof(t1)
```

```
[1] "list"
```

split into chars

```
u0 <- c("The quick brown fox jumps over the lazy dog")
u1 <- strsplit(u0, "")
u1
```

```
[[1]]
 [1] "T" "h" "e" " " "q" "u" "i" "c" "k" " " "b" "r" "o" "w" "n" " " "f" "o" "x"
[20] " " "j" "u" "m" "p" "s" " " "o" "v" "e" "r" " " "t" "h" "e" " " "l" "a" "z"
[39] "y" " " "d" "o" "g"
```

split into words

```
u2 <- strsplit(u0, " ")
u2
```

```
[[1]]
[1] "The"   "quick" "brown" "fox"   "jumps" "over"  "the"   "lazy"  "dog"
```

regex can be used in strsplit

```
v0 <- "all16i5need6is4a%long8vacation"
v1 <- strsplit(v0,split="[0-9,%]+")
v1
```

```
[[1]]
[1] "all"      "i"        "need"     "is"       "a"        "long"     "vacation"
```

---

2. paste()

```
a0 <- c("The quick brown fox jumps over the lazy dog")
a1 <- tolower(a0)
a1
```

```
[1] "the quick brown fox jumps over the lazy dog"
```