```r
# Character vector

# Character is a symbol in a written language like letters, punctuation, etc.
# String is a sequence of characters bound together.
# a lot of interesting data is in character form e.g., webpages, emails, surveys,...

# A character vector is a vector of text strings whose elements are specfied in quotes.

name1 <- c("Clinton", "Bush", "Obama");
name2 <- c("Reagan", "Carter", "Ford");
name <- c(name1,name2);
class(name)
rev.name <- rev(name);
paste(name, collapse = " ")                 # Combine the components of a vector
paste(name, collapse = "")                  # collapse: separated by the value of collapse

num <- nchar(name) # if we use "nchar" for a numeric vector, we can have the number of digits
num2 <- nchar(c("Bill Clinton", "Spider-Man"))
size <- length(name)   # Also works for a numeric and logical vectors.
size2 <- length(c("Bill Clinton", "Spider-Man"))

# Substring
phrase <- "Give me a break"
substr(phrase, 1, 4)
substr(phrase, nchar(phrase)-4, nchar(phrase))
substr(name, 1, 2)
substr(name, 1:6, 1:6)
substr(name, 1, 1:6)

# Replacement
substr(phrase, 1, 1) <- "L"
phrase
substr(phrase, 1, 4) <- "Show"
phrase

name
first.letters <- substr(name,1,1)
first.letters.scrambled <- sample(first.letters)
substr(name,1,1) <- first.letters.scrambled
name

# Converting other data types to strings
num_string <- as.character(num)
logi_string <- as.character(c(T,T,F,F,NA))
as.numeric(c(T,T,F,F,NA))

# Converting to lower or upper case
name_upper <- toupper(name1)
name_lower <- tolower(name2)

# If we want to remove " " in the output, we can use the cat function.
cat(c("Annie", "John", "Pam", "Sean", "Tom", "Ken"))
cat("What is \"R\"?")        # \ : escape character

# Built-in character vectors in R
LETTERS
letters
month.abb
month.name

# Find a long string of text from the web. The longer, the better. Copy and past this text below and
# save it as my.str.

Kennedy <- readLines("C:/Users/Min Seong Kim/Dropbox/R_programming/lecture2019/kennedy_speech.txt")
class(Kennedy)
```

```
str(Kennedy)
length(Kennedy)
my.str <- paste(Kennedy, collapse = " ")
my.str <- gsub(pattern = "[[:punct:]]", "", my.str)   #
http://www.endmemo.com/program/R/gsub.php

nchar(my.str)
my.list <- strsplit(my.str, split="")     # class of my.list is list
# strsplit(): split the elements of a character string
class(my.list)
my.chars <- unlist(my.list)                # We want to work with character vector

class(my.char)

my.chars_lower <- tolower(my.chars)
my.chars <- my.chars[my.chars %in% c(letters)]
my.chartab <- table(my.chars)
my.chartab
plot(my.chartab, xlab="Characters", ylab="Counts")


# Exercises
################################################################################

presidents <- c("Bill Clinton", "George Bush", "Ronald Reagan", "Jimmy Carter", "Gerald
Ford")
# Split each string in the vector presidents into two strings, according to the first
and last names
# Save the results as "president.list"
# create president.first as a vector of president.

# Substrings, replacement
bill <- "Bill Clinton"
others.first <- c("George", "Ronald", "Jimmy", "Gerald")
others.second <- c("Bush", "Reagan", "Carter", "Ford")
# Using paste combine others.first and others.second to have others
# Display the president names that include "al" (hint: use grep() function)
others <- paste(others.first, others.second)
grep("al",others,value=TRUE)

# Display the last 5 characters from each of the strings in the vector 'others'
# Combine 'bill' and 'others' into one vector 'president' and replace the first 3
letters in each
# strings in 'president' with 'abc'
```