Assignment Week 6

Jihun Bae

Assignment 6.1.1

Fill in the black below to produce the outcome.

1. Q6.1.1.1

```
## [1] ".a.b.c"
```

2. Q6.1.1.2

```
## [1] "$^$"
```

Assignment 6.1.2

Using the corpus of common words in stringr::words, create regular expressions to find all words that meet the following questions.

*Use str_subset(stringr::words,)

Questions

1. Find all words that start with a vowel.
2. Find all words that only contain consonants. (*Hint: thinking about matching "not"-vowels.)
3. Find all words that end with "ed", but not with "eed". (e.g., "ed", "bed", "hundred", "red")
4. Find all words that end with "ing" or "ise".

Answers 6.1.2

1. Q6.1.2.1

```
##   [1] "a"         "able"      "about"     "absolute"  "accept"
##   [6] "account"   "achieve"   "across"    "act"       "active"
##  [11] "actual"    "add"       "address"   "admit"     "advertise"
##  [16] "affect"    "afford"    "after"     "afternoon" "again"
##  [21] "against"   "age"       "agent"     "ago"       "agree"
##  [26] "air"       "all"       "allow"     "almost"    "along"
##  [31] "already"   "alright"   "also"      "although"  "always"
##  [36] "america"   "amount"    "and"       "another"   "answer"
##  [41] "any"       "apart"     "apparent"  "appear"    "apply"
```

```
##  [46] "appoint"    "approach"     "appropriate" "area"        "argue"
##  [51] "arm"        "around"       "arrange"     "art"         "as"
##  [56] "ask"        "associate"    "assume"      "at"          "attend"
##  [61] "authority"  "available"    "aware"       "away"        "awful"
##  [66] "each"       "early"        "east"        "easy"        "eat"
##  [71] "economy"    "educate"      "effect"      "egg"         "eight"
##  [76] "either"     "elect"        "electric"    "eleven"      "else"
##  [81] "employ"     "encourage"    "end"         "engine"      "english"
##  [86] "enjoy"      "enough"       "enter"       "environment" "equal"
##  [91] "especial"   "europe"       "even"        "evening"     "ever"
##  [96] "every"      "evidence"     "exact"       "example"     "except"
## [101] "excuse"     "exercise"     "exist"       "expect"      "expense"
## [106] "experience" "explain"      "express"     "extra"       "eye"
## [111] "idea"       "identify"     "if"          "imagine"     "important"
## [116] "improve"    "in"           "include"     "income"      "increase"
## [121] "indeed"     "individual"   "industry"    "inform"      "inside"
## [126] "instead"    "insure"       "interest"    "into"        "introduce"
## [131] "invest"     "involve"      "issue"       "it"          "item"
## [136] "obvious"    "occasion"     "odd"         "of"          "off"
## [141] "offer"      "office"       "often"       "okay"        "old"
## [146] "on"         "once"         "one"         "only"        "open"
## [151] "operate"    "opportunity"  "oppose"      "or"          "order"
## [156] "organize"   "original"     "other"       "otherwise"   "ought"
## [161] "out"        "over"         "own"         "under"       "understand"
## [166] "union"      "unit"         "unite"       "university"  "unless"
## [171] "until"      "up"           "upon"        "use"         "usual"
```

  2. Q6.1.2.2

```
## [1] "by"   "dry" "fly" "mrs" "try" "why"
```

  3. Q6.1.2.3

```
## [1] "bed"      "hundred" "red"
```

  4. Q6.1.2.4

```
##  [1] "advertise" "bring"     "during"    "evening"   "exercise"  "king"
##  [7] "meaning"   "morning"   "otherwise" "practise"  "raise"     "realise"
## [13] "ring"      "rise"      "sing"      "surprise"  "thing"
```

Assignment 6.1.3

Create a regular expression that will match the following telephone numbers.

"123-456-7890", "(123)456-7890", "(123) 456-7890", "(123)5-2351"

Questions

  1. match telephone number "123-456-7890"
  2. match telephone numbers "(123)456-7890" and "(123) 456-7890"

Answers 6.1.3

1. Q6.1.3.1

```
## [1] "123-456-7890"
```

2. Q6.1.3.2

```
## [1] "(123)456-7890"  "(123) 456-7890"
```

Assignment 6.2

Consider the following dataset for healthcare organizations (hcodata): How can you find the organizations with limited liability? (Hint: you should find the name of organization includes "B.V." (or equivalent).

Answers 6.2

```
## [1] "ContactZorg BV"      "BAVEL & Partners B.V." "Handenzorg (bv)"
## [4] "O'Thuiszorg BV"      "MOHSA b.V"             "Stichting bv."
```

Assignment 6.3

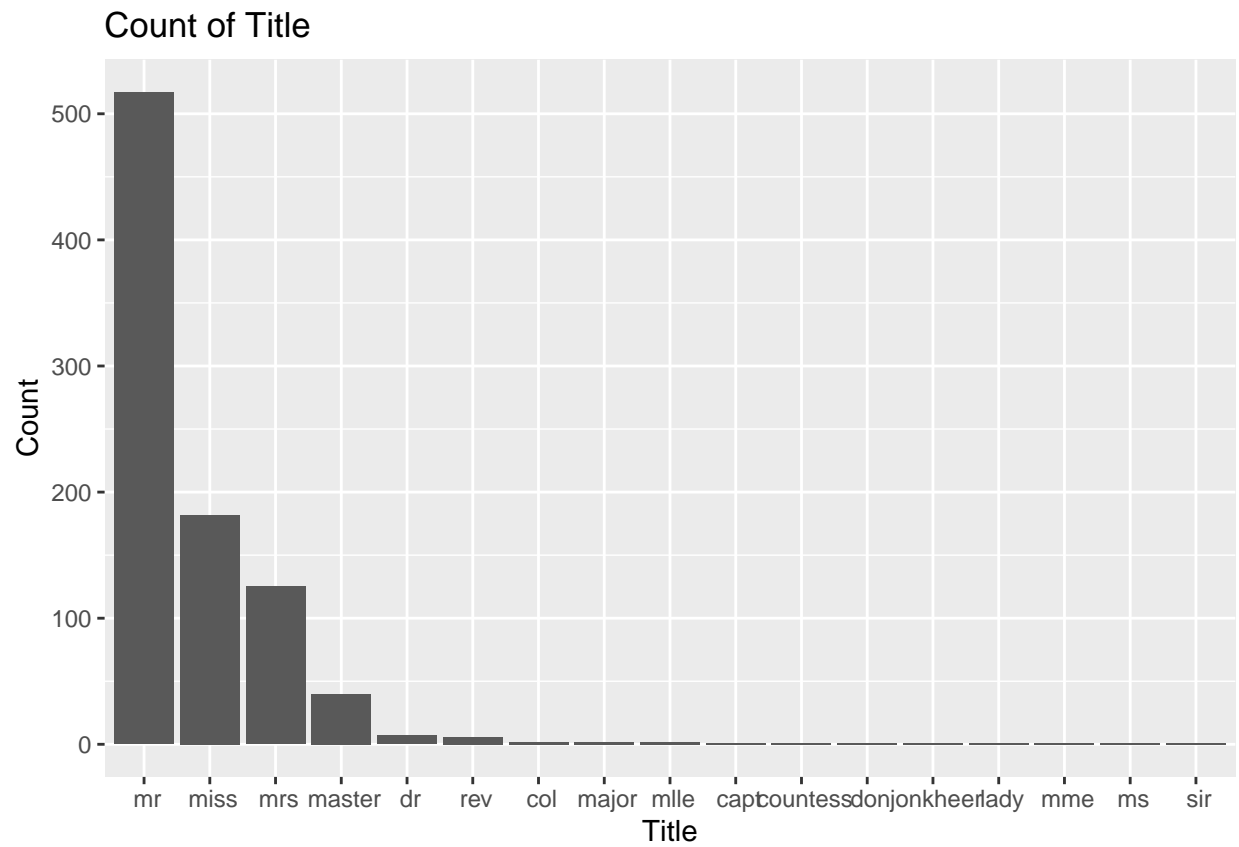Extract and plot features from "titanic_train" data.

Questions

1. Extract titles from "Names". View the first 10 rows of titles using head() function.
2. Plot the new variable "title" using ggplot2.
3. Extract Family names from "Names". View the first 10 rows of Family names using head() function.
4. Plot the 10 most frequent family names using ggplot2.

Answers 6.3

1. Q6.3.1

```
##  [1] " mr"     " mrs"    " miss"   " mrs"    " mr"     " mr"     " mr"
##  [8] " master" " mrs"    " mrs"
```

2. Q6.3.2

Count of Title

3. Q6.3.3

```
##  [1] "braund"    "cumings"    "heikkinen" "futrelle" "allen"     "moran"
##  [7] "mccarthy"  "palsson"    "johnson"   "nasser"
```

4. Q6.3.4

Count of Family Name