STAT 133 Fall 12
Section Worksheet: Regular Expressions

Write down a regular expression to match the following:

Words with @ symbols in them, e.g., vi@gra

```
"\\<[[:alpha:]]+@[[:alpha:]]+\\>"
```

An ip address (4 sets of 1-3 digits separated by periods, e.g., 124.32.6.240)

```
"[0-9]{1,3}\\.[0-9]{1,3}\\.[0-9]{1,3}\\.[0-9]{1,3}"
```

A typical email address that ends with .com, .edu, .net, .org, or .gov

```
"[-[:alnum:]]+@[[:alnum:]._]+\\.(com|edu|net|gov)"
```

Consider the following character vector,

```
> movies
[1] "The Shawshank Redemption (1994)"
[2] "The Godfather (1972)"
[3] "The Godfather: Part II (1974)"
[4] "Pulp Fiction (1994)"
[5] "The Good, the Bad and the Ugly (1966)"
[6] "12 Angry Men (1957)"
```

What is the return expression from each of the following function calls:

```
> grep("I{2,}", movies)

3

> grep("Go+d", movies)

2 3 5

> gregexpr("\\(.*\\)", movies[1])

26 length 6

> gsub("[0-9]", "", movies[6])

" Angry Men ()"

> gsub("[[:blank:]].*$", "", movies[5])

"The"

> gsub(" \\(.*$", "", movies[5])

"The Good, the Bad and the Ugly"
```