

# Tutorial 12 - Problem 2

*Chelsea Weibel*

*November 17, 2017*

```
library(stringr)
```

## a. Time after noon, but before midnight (12:00-23:59)

```
time=c('1500','23:59','12:40','01:40','Cow')

print(str_detect(time,'(1[2-9]|2[0-3]):[0-5][0-9]'))
```

```
## [1] FALSE TRUE TRUE FALSE FALSE
```

This expression will select all times reported in the afternoon (from 12:00-23:59) because the expression reads as follows:

- \* 1[2-9] | 2[0-3] = there can be a 1 in the first position if there is a 2-9 in the second position (i.e. hours 12:00-19:00) OR there can be a 2 in the first position if there is a 0-3 in the second position (i.e. hours 20:00-23:00)
- \* : = there will be a colon in the third position
- \* [0-5] = there will be a 0-5 in the fourth (minute) position
- \* [0-9] = there will be any digit in the fifth (seconds) position

## b. Genus species names expressed in the format G. species

```
species=c('P. cynocephalus','Dog','C. lupus','H. sapiens','Cow')

print(str_detect(species,'([A-Z][.] [a-z]+)'))
```

```
## [1] TRUE FALSE TRUE TRUE FALSE
```

This expression will select only genus species names in the format G. species because it reads as follows:

- \* [A-Z] = the first letter must be a capital
- \* [.] = a period must follow in the second position
- \* Followed by a space in the third position
- \* [a-z]+ = there will be at least one lower case letter followed by any number of lowercase letters

## c. Social security numbers (###-##-####)

```
SSN=c('111-11-1111','555-555-4285','919-19-1919','56','Cow')
```

```
print(str_detect(SSN,'[0-9]{3}-[0-9]{2}-[0-9]{4}'))
```

```
## [1] TRUE FALSE TRUE FALSE FALSE
```

This expression will select for all social security numbers because it reads as follows:

- \* [0-9]{3} = there must be three of any number

- \* - = followed by a dash

- \* [0-9]{2} = then two of any number

- \* - = followed by another dash

- \* [0-9]{4} = then four of any number