General Notes

- You will submit a minimum of three files, the core files must conform to the following naming conventions (including capitalization and underscores). 123456789 is a placeholder, please replace these nine digits with your nine-digit Bruin ID. The files you must submit are:

  1. *123456789_stats102a_hw3.R*: An R script file containing all of the functions you wrote for the homework. The first line of your .Rmd file, **after loading libraries, should be sourcing this script file**.
  2. *123456789_stats102a_hw3.Rmd*: Your markdown file which generates the output file of your submission.
  3. *123456789_stats102a_hw3.html/pdf*: Your output file, either a PDF or an HTML file depending on the output you choose to generate.
  4. *included image files*: You may name these what you choose, but you must include all of the image files you generated for your structured flowcharts, otherwise your file will not knit. One way to attach an image file is to use ![*title*](*Images/example1.png*).

  If you fail to submit any of the required core files you will receive ZERO points for the assignment. If you submit any files which do not conform to the specified naming convention, you will receive half credit for the assignment.

- **Your .Rmd file must be knittable. If your .Rmd file does not knit in graders' computer, you will receive at most half credit for the assignment.**
  The two most common reason files fail to knit are because of workspace/directory structure issues and because of missing include files. To remedy the first, ensure all of the file paths in your document are relative paths pointing at the current working directory. To remedy the second, simply make sure you upload any and all files you source or include in your .Rmd file.

- Your coding should adhere to the tidyverse style guide: https://style.tidyverse.org/.

- Any functions/classes you write should have the corresponding comments as the following format.

```
my_function <- function(x, y, ...){
#A short description of the function
#Args:
#x: Variable type and dimension
#y: Variable type and dimension
#Return:
#Variable type and dimension
Your codes begin here
}
```

NOTE: *Everything* you need to do this assignment is here, in your class notes, or was covered in discussion or lecture.

- **DO NOT** look for solutions online.
- **DO NOT** collaborate with anyone inside (or outside) of this class.
- Work **INDEPENDENTLY** on this assignment.
- **EVERYTHING** you submit **MUST** be 100% your, original, work product. Any student suspected of plagiarizing, in whole or in part, any portion of this assignment, will be **immediately** referred to the Dean of Student's office without warning.

Homework Requirements You are going to produce several examples of regex.

You are trying to get the **lowest** score you can using the scoring function provided below. The scoring works as follows, you will get:

1. 10 points for every word incorrectly matched or not matched.
2. 1 point for every character in your regex expression.

```
library(stringr)
regex_golf <- function(x, y, regex) {
  xmatch <- str_extract_all(x, regex) == x
  matched_x <- x[xmatch]
  unmatched_x <- x[!xmatch]
  ymatch <- str_extract_all(y, regex) == y
  matched_y <- y[ymatch]
  unmatched_y <- y[!ymatch]
  penalty <- 10 * sum(!xmatch, ymatch)
  score <- nchar(regex) + penalty
  invisible(list( score = score, matched_x = matched_x,
            unmatched_x = unmatched_x, matched_y = matched_y,
            unmatched_y = unmatched_y
))
}
```

For example, if we had the following two character vectors x and y. The goal is to create a regular

x	y
Ella	Emma
Mila	Olivia
Isla	

expression pattern which will match all of the values in x and none of the values in y.

If we attempted the pattern `(El|Mi)la` as our regular expression, we would get the following results:

Matched x	Unmatched x	Matched y	Unmatched y
Ella	Isla		Emma
Mila			Olivia

for a score of 19 (a 10 point penalty for not matching **Isla** and 9 points for the regular expression being 9 characters long).

A better attempt could be the pattern \\w{2}la as our regular expression, we would get the following results:

Matched x	Unmatched x	Matched y	Unmatched y
Ella			Emma
Mila			Olivia
Isla			

for a score of 7

1: Regex Golf

Write regex patterns to play Regex Golf on the following lists. For each pattern, store the pattern as a variable named in the following way **pat_x** where the x is replaced by the number of the exercise you are doing. You should save these patterns in your **.R** file.

In your output file, for each problem, report your score, along with displaying the strings in x you failed to match and the strings in y you did match.

All of the strings you will need to match or not are in the files wordlists.RDS and wordlists.RData on CCLE (use whichever you prefer).

Please choose 10 of the following 12 problems to do. You are NOT expected to do all of these perfectly. In fact, on some of the more challenging ones, it is possible you will struggle to find a pattern which selects every string in the x list. THIS IS OKAY. The expectation is you will, in this order:

1. Have fun
2. Try hard
3. Do your best

These are openly available problems which do have solutions available online.

DO NOT search these out. That is not the point of this exercise. Rather, the point is for you to play with regular expressions to get a better feel for them and to flex your creative thinking muscles.

Any solution you submit you will be expected to be able to explain. If you do not understand your solution well enough to explain it, do not submit it.

1. **Warmup**

	x	y
1	afoot	Atlas
2	catfoot	Aymoro
3	dogfoot	Iberic
4	fanfoot	Mahran
5	foody	Ormazd
6	foolery	Silipan
7	foolish	altared
8	fooster	chandoo
9	footage	crenel
10	foothot	crooked
11	footle	fardo
12	footpad	folksy
13	footway	forest
14	hotfoot	hebamic
15	jawfoot	idgah
16	mafoo	manlike
17	nonfood	marly
18	padfoot	palazzi
19	prefool	sixfold
20	sfoot	tarrock
21	unfool	unfold

2. **Anchors**

	x	y
1	Mick	Kickapoo
2	Rick	Nickneven
3	allocochick	Rickettsiales
4	backtrick	billsticker
5	bestick	borickite
6	candlestick	chickell
7	counterprick	fickleness
8	heartsick	finickily
9	lampwick	kilbrickenite
10	lick	lickpenny
11	lungsick	mispickel
12	potstick	quickfoot
13	quick	quickhatch
14	rampick	ricksha
15	rebrick	rollicking
16	relick	slapsticky
17	seasick	snickdrawing
18	slick	sunstricken
19	tick	tricklingly
20	unsick	unlicked
21	upstick	unnickeled

3. **Ranges**

	x	y
1	abac	beam
2	accede	buoy
3	adead	canjac
4	babe	chymia
5	bead	corah
6	bebed	cupula
7	bedad	griece
8	bedded	hafter
9	bedead	idic
10	bedeaf	lucy
11	caba	martyr
12	caffa	matron
13	dace	messrs
14	dade	mucose
15	daff	relose
16	dead	sonly
17	deed	tegua
18	deface	threap
19	faded	towned
20	faff	widish
21	feed	yite

4. **Backrefs**

	x	y
1	allochirally	anticker
2	anticovenanting	corundum
3	barbary	crabcatcher
4	calelectrical	damnably
5	entablement	foxtailed
6	ethanethiol	galvanotactic
7	froufrou	gummage
8	furfuryl	gurniad
9	galagala	hypergoddess
10	heavyheaded	kashga
11	linguatuline	nonimitative
12	mathematic	parsonage
13	monoammonium	pouchlike
14	perpera	presumptuously
15	photophonic	pylar
16	purpuraceous	rachioparalysis
17	salpingonasal	scherzando
18	testes	swayed
19	trisectrix	unbridledness
20	undergrounder	unupbraidingly
21	untaunted	wellside

5. **Abba**

	x	y
1	acritan	abba
2	aesthophysiology	anallagmatic
3	amphimictical	bassarisk
4	baruria	chorioallantois
5	calomorphic	coccomyces
6	disarmature	commotive
7	effusive	engrammatic
8	fluted	glossoscopia
9	fusoid	hexacoralla
10	goblinize	hippogriffin
11	nihilistic	inflammableness
12	noisefully	otto
13	picrorhiza	overattached
14	postarytenoid	saffarid
15	revolutionize	sarraceniaceae
16	suprasphanoidal	scillipicrin
17	suspenseful	tlapallan
18	tapachula	trillion
19	transmit	unclassably
20	unversatile	unfitting
21	vibetoite	unsmelled
22		warrandice

6. **A man, a plan**

	x	y
1	civic	arrogatingly
2	deedeed	camshach
3	degged	cinnabar
4	hallah	defendress
5	kakkak	derivedly
6	kook	gourmet
7	level	hamleteer
8	murdrum	hydroaviation
9	noon	lophine
10	redder	nonalcohol
11	repaper	outslink
12	retter	pretest
13	reviver	psalterium
14	rotator	psorosperm
15	sexes	scrummage
16	sooloos	sporous
17	tebbet	springer
18	tenet	sunburn
19	terret	teleoptile
20		unstuttering
21		womanways

7. **Prime**

	x	y
1	xx	xxxx
2	xxx	xxxxxx
3	xxxxx	xxxxxxxx
4	xxxxxxx	xxxxxxxxx
5	xxxxxxxxxxx	xxxxxxxxxx
6	xxxxxxxxxxxxx	xxxxxxxxxxxx
7	xxxxxxxxxxxxxxxx	xxxxxxxxxxxxxx
8	xxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxx
9	xxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxx
10	xxxxxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxx
11	xxxxxxxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxx
12	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxx
13	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxx
14	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxxx
15	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxxx
16	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxxxx
17	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxxxx
18	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxxxx
19	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxxxxx
20	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxxxxx

8. **Four**

	x	y
1	Makaraka	Ludgate
2	Wasagara	Mitsukurinidae
3	degenerescence	Ternstroemiaceae
4	desilicification	arrhythmical
5	elevener	bleater
6	hipponosological	energetics
7	homoeomorphy	inthrow
8	homologous	mecopterous
9	ileocolotomy	multum
10	intervisibility	naphthalene
11	jararaca	nullibicity
12	locomotory	observancy
13	micropoikilitic	overpunishment
14	odontonosology	overregularly
15	parhomologous	overwilily
16	pogonotomy	participator
17	promonopolist	predisable
18	protohomo	reyield
19	pseudoprimitivism	rubeola
20	tocororo	traitorlike
21	unintelligibility	unregainable

9. **Order**

	x	y
1	access	analyse
2	accloy	balanism
3	adeem	baronet
4	aflow	biddable
5	aglow	griefless
6	beefin	harebrain
7	befist	jestword
8	billot	laicize
9	bossy	marvelry
10	certy	oriole
11	chintz	pickietar
12	chips	preferee
13	chort	primness
14	cloop	pulghere
15	coost	rebirth
16	demos	scupper
17	fitty	serigraph
18	flory	sororize
19	flossy	theowman
20	ghost	unfrayed
21	mopsy	wagonman

10. **Triples**

	x	y
1	000000000	000000005
2	000000003	000000008
3	000000006	000000010
4	000000009	000000011
5	000000012	000000014
6	000000015	018990130
7	066990060	112057285
8	140091876	159747125
9	173655750	176950268
10	312440187	259108903
11	321769005	333162608
12	368542278	388401457
13	390259104	477848777
14	402223947	478621693
15	443512431	531683939
16	714541758	704168662
17	747289572	759282218
18	819148602	769340942
19	878531775	851936815
20	905586303	973816159
21	953734824	979204403

11. **Glob**

	x	y
1	*err* matches superreform	*anapaestical* matches anapaestically
2	*falle*ess matches unfallenness	*chegonio matches archegoniophore
3	*il*log* matches unphilological	*dissoluti* matches dissolutional
4	*plen*tud* matches overplenitude	*domestica matches domesticality
5	*taiodi* matches pentaiodide	*expedition matches expedition
6	*viceberry matches serviceberry	*hormog matches hormogonium
7	bowdl* matches bowdlerism	*stipular* matches infrastipular
8	bron*hopleur*sy matches bronchopleurisy	*strabis matches strabismal
9	chromatophobia matches chromatophobia	cathartica matches cathartically
10	cockneyla* matches cockneyland	di matches gerundively
11	colorlessly matches colorlessly	hacean matches zoanthacean
12	cretefaction matches cretefaction	headmist matches headmistress
13	downrightly matches downrightly	herwi matches trencherwise
14	leather* matches leatherbark	iemphraxia matches cardiemphraxia
15	mitogenet* matches mitogenetic	kmak matches packmaking
16	palindrom* matches palindromic	mbable* matches unclimbable
17	parallelepiped matches parallelepiped	nspi*tor matches inspirator
18	primigenial matches primigenial	ocumidi matches pseudocumidine
19	puppe* matches puppetlike	raretinal* matches intraretinal
20	resurrender matches resurrender	tte matches whitterick
21	wreathwi* matches wreathwise	uefoliate matches quinquefoliate

12. **Alphabetical**

	x	y
1	aerate aerate arrest errant serene tanner testes	aerate astern assess enseat senate street tsetse
2	aerate assent assent assert rester retest tenant	aerate rennet errant enseat rerent senate testes
3	aerate assert rearer renter resent serene teaser	arrest assess assess assent astern searer testes
4	aerate easter easter tenant tester testes tsetse	assert assess errata enseat earner seater serene
5	arrest arrest easter entree errant resent senate	assert astern staree senate snarer tanner tester
6	assent assess assets estate resent staree teaser	assert strata rerent rerent tanner testes tsetse
7	assert astern renter rerent resent staree street	assess easter entree rester reseat seater tartar
8	assert enseat entree errata rennet teaser tsetse	astern assets rearer rearer assess rearer testes
9	assert rennet renter reseat rester serene tenant	astern easter taster serene reseat taster tester
10	assess easter estate rennet rennet tenant testes	earner entree rerent reseat teaser strata staree
11	assess easter estate rerent resent retest snarer	earner errant estate taster reseat estate taster
12	assess renter renter searer seater snarer testes	enseat astern arrest enseat searer seater tenant
13	astern enseat entree serene staree tartar tartar	errant errant senate renter rearer street tsetse
14	astern rennet retest searer snarer tartar tester	rennet rennet assent errant rester staree tester
15	enseat errata seater senate strata teaser tsetse	rennet snarer senate retest tanner tartar tsetse
16	entree searer staree taster taster tenant testes	retest astern arrest tsetse strata senate tsetse
17	rerent rester tanner tartar teaser teaser testes	searer errant teaser staree assess teaser tsetse