Problem set 1:

1)

SortedMinus2<-Countsminus2[order(-Countsminus2[,"matches"]),]

SortedMinus2[1:20,1:2]

string matches

10607 La 2127

12253 Ce 1690

10892 Ral 536

11694 Ther 535

3360 Ace 521

10574 Av 505

10782 Pi 457

11726 Fer 404

12505 Fer 404

10647 Alc 199

10267 Boid 112

10969 Sco 103

11387 Such 102

15001 Calcar 75

15439 Silic 65

5611 Articula 51

14492 Articula 51

3894 Ostraco 42

4617 Aca 40

4938 Producti 38

Lari (suborder that includes gulls)

Cete (clade that comprises cetacea)

Ralli (??????)

Theria (non-monotreme mammals)

Acera (Gastropods) \*\*\*\*\*

Aves (birds)

Pici (woodpeckers and related taxa)

Ferae (carnivorous mammals and other shtuff)

Alcae (Puffins!!)

Boidae ( constrictor snakes)

Scopi (shorebirds?)

Suchia ( crocodilians and other archosaurs)

Calcarea (calcareous sponges) \*\*\*\*\*

Silicea (siliceous sponges) \*\*\*\*\*

Articulata (ambiguous--crinoids, articulated brachiopods, or bryozoans)\*\*\*\*

Ostracoda (benthic shrimpy things) \*\*\*\*\*?

Acari (arachnids)

Productida (brachiopods)\*\*\*\*\*

Taxon names designated with a “\*\*\*\*” are likely to be pyritized because they are epifaunal or infaunal and therefore are more likely to be preserved in anoxic conditions.

my\_taxa<-c("infaunal","epifaunal","anoxic","euxinic","silic","calc","brachiopod","ammon","trilobi","crinoid","foram","shell","spicul")

my\_counts<-count\_uses(my\_taxa,0,pyr)

my\_counts[order(-my\_counts[,"matches"]),]

string matches

6 calc 269

5 silic 233

11 foram 232

3 anoxic 131

12 shell 110

8 ammon 77

4 euxinic 69

9 trilobi 31

7 brachiopod 24

13 spicul 23

1 infaunal 12

10 crinoid 10

2 epifaunal 3

I chose some taxa that I know from past classes that are commonly preserved as pyrite (due to their benthic nature or hard-shell biology), along with some terms that describe those qualities (like chitin and anoxic). There were fewer matches overall, largely because I truncated some terms myself (like “foram” and “calc”) and others need no truncation (“epifaunal”). I trust my list more, because I was able to determine how much I wanted each term truncated—“calc” could give “calcareous”, “calcifiers”, “calcite”, etc., but if I were starting from “calcareous” I would need to truncate it by six letters, which would obliterate shorter terms. So, my own compiled list is tailored to get the results I want.