

# Lab: Week 3

36-350 – Statistical Computing

Week 3 – Fall 2020

Name: Kimberly Zhang

Andrew ID: kyz

You must submit **your own** lab as a PDF file on Gradescope.

---

## Basic Strings

### Question 1

(4 points)

Notes 3A (4)

Make a single string with your first and last names, and (programmatically) count the number of characters.

```
my_name = "Kimberly Zhang"
nchar(my_name)
```

```
## [1] 14
```

### Question 2

(4 points)

Notes 3A (4)

Make a vector of strings of length two, with your first name first and then last name, and (programmatically) count the number of characters in each.

```
my_name = c("Kimberly", "Zhang")
nchar(my_name)
```

```
## [1] 8 5
```

### Question 3

(4 points)

Refer Back to Week 1

Output the vector with the second element removed. Do this *two* different ways: by indicating which element you want, in square brackets; and by indicating which element you *don't* want, in square brackets with a minus sign.

```
match_first = which(my_name == "Kimberly")
match_second = which(my_name == "Zhang")
my_name[match_first]
```

```
## [1] "Kimberly"
```

```
my_name[-match_second]
```

```
## [1] "Kimberly"
```

## Question 4

(4 points)

Notes 3A (5)

Transform the vector so that the words have all uppercase characters.

```
toupper(my_name)
```

```
## [1] "KIMBERLY" "ZHANG"
```

## Question 5

(4 points)

Not Covered in Notes

Make a string with “Regards,” and then your first name, separated by a newline character. Print it to the console via `print()` and then via `cat()`. Comment on the difference.

```
signature = "Regards,\nKimberly"
print(signature)
```

```
## [1] "Regards,\nKimberly"
```

```
cat(signature)
```

```
## Regards,
## Kimberly
```

`print` shows literal source code output whereas `cat` converts the special characters into readable form.

## Question 6

(4 points)

Notes 3A (4)

Count the number of characters in the string in the last question. Is the comma counted? The newline counted?

```
nchar(signature)
```

```
## [1] 17
```

Both the comma and the newline character were counted.

---

```
bill = "Bill Clinton"
others = c("George Bush", "Ronald Reagan", "Jimmy Carter", "Gerald Ford")
```

---

## Question 7

(4 points)

Notes 3A (6)

Display the first five characters of the string `bill`.

```
substr(bill, 1, 5)
```

```
## [1] "Bill "
```

## Question 8

(4 points)

Refer Back to Week 1

Join `bill` and `others` together into one vector of strings, called `presidents`.

```
presidents = append(others, bill)
presidents
```

```
## [1] "George Bush" "Ronald Reagan" "Jimmy Carter" "Gerald Ford"
## [5] "Bill Clinton"
```

## Question 9

(4 points)

Notes 3A (6)

Display the last three letters of each of the strings in the vector `presidents`.

```
substr(presidents, nchar(presidents)-2, nchar(presidents))
```

```
## [1] "ush" "gan" "ter" "ord" "ton"
```

## Regular Expressions + Extraction

Below, we read in lines from Hillary Clinton's speech at the 2016 Democratic National Convention. (The `readLines()` function reads in a text file line-by-line: each line is a character string, and the vector of strings makes up the entire file.)

```
clinton.lines = readLines("http://www.stat.cmu.edu/~pfreeman/clinton.txt")
```

## Question 10

(4 points)

Notes 3B (2-3)

Using a regex literal and `grep()`, display the line numbers in `clinton.lines` that contain "Bill". Also, display the lines of text that contain "Bill". (Hint: try setting `value=FALSE` and `value=TRUE`.)

```
grep("Bill", clinton.lines, value = FALSE)
```

```
## [1] 4 9
```

```
grep("Bill", clinton.lines, value = TRUE)
```

```
## [1] "And Bill, that conversation we started in the law library 45 years ago
... it is still going strong."
## [2] "We heard the man from Hope, Bill Clinton, and the man of hope, Barack
Obama. America is stronger because of President Obama's leadership. And I'm
better because of his friendship."
```

## Question 11

(4 points)

*Notes 3B (3) and Refer Back to Week 1*

Show, using the line numbers from the last problem from `grep()` with `value=FALSE`, how the lines of text containing “Bill” can be retrieved from `clinton.lines` through direct indexing, i.e., without a call to `grep()` with `value=TRUE`. Note: this is just to check your understanding of indexing; when it comes to retrieving and displaying matched text, using `grep()` with `value=TRUE` is easier; you should stick to it as the default, below.

```
line_num = grep("Bill", clinton.lines)
clinton.lines[line_num]
```

```
## [1] "And Bill, that conversation we started in the law library 45 years ago
... it is still going strong."
## [2] "We heard the man from Hope, Bill Clinton, and the man of hope, Barack
Obama. America is stronger because of President Obama's leadership. And I'm
better because of his friendship."
```

## Question 12

(4 points)

*Notes 3B (2-3)*

Display the lines of text in `clinton.lines` that contain “Bill” or “Chelsea”. Also, display the lines of text that contain “Bill” or “Chelsea” or “Tim”.

```
grep("Bill|Chelsea", clinton.lines, value = TRUE)
```

```
## [1] "And Chelsea, thank you. I am so proud to be your mother and so proud of
the woman you've become. Thank you for bringing Mark into our family and
Charlotte and Aidan into the world."
## [2] "And Bill, that conversation we started in the law library 45 years ago
... it is still going strong."
## [3] "We heard the man from Hope, Bill Clinton, and the man of hope, Barack
Obama. America is stronger because of President Obama's leadership. And I'm
better because of his friendship."
```

```
grep("Bill|Chelsea|Tim", clinton.lines, value = TRUE)
```

```
## [1] "And Chelsea, thank you. I am so proud to be your mother and so proud of
the woman you've become. Thank you for bringing Mark into our family and
Charlotte and Aidan into the world."
## [2] "And Bill, that conversation we started in the law library 45 years ago
... it is still going strong."
## [3] "We heard the man from Hope, Bill Clinton, and the man of hope, Barack
Obama. America is stronger because of President Obama's leadership. And I'm
better because of his friendship."
```

```
## [4] "And for those of you out there who are just getting to know Tim Kaine you will soon understand why the people of Virginia keep promoting him from city council and mayor to governor and now senator. And he will make our whole country proud as our vice president."
```

```
## [5] "Well, I've had the privilege to work closely with our troops and our veterans for many years, including as a senator on the Armed Services Committee, and I know how wrong he is. Our military is a national treasure. We entrust our commander in chief to make the hardest decisions our nation faces, decisions about war and peace, life and death. A president should respect the men and women who risk their lives to serve our country...including Captain Khan and the sons of Tim Kaine and Mike Pence, both Marines."
```

## Question 13

*(4 points)*

*Notes 3B (4-5)*

Display the lines of text in `clinton.lines` that contain "Trump" and then a punctuation mark.

```
grep("Trump[[:punct:]]", clinton.lines, value = TRUE)
```

```
## [1] "Well, we heard Donald Trump's answer last week at his convention. He wants to divide us from the rest of the world and from each other. He's betting that the perils of today's world will blind us to its unlimited promise. He's taken the Republican Party a long way, from morning in America to midnight in America."
```

```
## [2] "Yes, those were actually Donald Trump's words in Cleveland. And they should set off alarm bells for all of us. Really? I alone can fix it? Isn't he forgetting troops on the front lines, police officers and firefighters who run toward danger, doctors and nurses who care for us, teachers who change lives, entrepreneurs who see possibilities in every problem, mothers who lost children to violence and are building a movement to keep other kids safe? He's forgetting every last one of us."
```

```
## [3] "But then I also imagine people are thinking out there, but Trump, he's a businessman, he must know something about the economy."
```

```
## [4] "Now, Donald Trump, Donald Trump says, and this is a quote, \"I know more about ISIS than the generals do.\" No, Donald, you don't."
```

```
## [5] "And we will stand up against mean and divisive rhetoric wherever it comes from. You know, for the past year many people made the mistake of laughing off Donald Trump's comments, excusing him as an entertainer just putting on a show."
```

```
## [6] "Now, at first, at first, I admit, I couldn't he meant it either. It was just too hard to fathom that someone who wants to lead our nation could say those things, could be like that. But here's the sad truth: There is no other Donald Trump, this is it."
```

```
## [7] "So enough with the bigotry and the bombast. Donald Trump's not offering real change, he's offering empty promises. And what are we offering? A bold agenda to improve the lives of people across our country to keep you safe, to get you good jobs, to get your kids the opportunities they deserve."
```

## Question 14

*(4 points)*

*Notes 3B (4-5)*

Display the lines of text in `clinton.lines` that contain “Trump” and then anything but a punctuation mark.

```
grep("Trump[^[:punct:]]", clinton.lines, value = TRUE)
```

```
## [1] "He wants us to fear the future and fear each other. Well, you know, a
great Democratic President Franklin Delano Roosevelt came up with the perfect
rebuttal to Trump more than 80 years ago during a much more perilous time: The
only thing we have to fear is fear itself!"
## [2] "Now, you didn't hear any of this, did you, from Donald Trump at his
convention? He spoke for 70-odd minutes-and I do mean odd-and he offered zero
solutions. But we already know he doesn't believe these things. No wonder he
doesn't like talking about his plans. You might have noticed I love talking
about mine."
## [3] "It's just not right that Donald Trump can ignore his debts and students
and families can't refinance their debts."
## [4] "Well, let's take a closer look, shall we? In Atlantic City, 60 miles
from here, you will find contractors and small businesses who lost everything
because Donald Trump refused to pay his bills."
## [5] "But think of this. People who did the work and needed the money, not
because he couldn't pay them, but because he wouldn't pay them. He just stiffed
them. And you know that sales pitch he's making to be president, put your faith
in him and you'll win big? That's the same sales pitch he made to all those
small businesses. Then Trump walked away and left working people holding the
bag."
## [6] "He also talks a big game about putting America first. Well, please
explain what part of \"America first\" leads him to make Trump ties in China,
not Colorado, Trump suits in Mexico, not Michigan, Trump furniture in Turkey,
not Ohio, Trump picture frames in India, not Wisconsin?"
## [7] "Donald Trump says he wants to make America great again. Well, he could
start by actually making things in America again."
## [8] "Now, Donald Trump, Donald Trump says, and this is a quote, \"I know
more about ISIS than the generals do.\" No, Donald, you don't."
## [9] "So just ask yourself, do you really think Donald Trump has the
temperament to be commander in chief? Donald Trump can't even handle the rough
and tumble of a presidential campaign. He loses his cool at the slightest
provocation, when he's gotten a tough question from a reporter, when he's
challenged in a debate, when he sees a protester at a rally. Imagine, if you
dare, imagine, imagine him in the Oval Office facing a real crisis."
## [10] "And in the end, it comes down to what Donald Trump doesn't get:
America is great because America is good!"
```

## Question 15

*(4 points)*

*Notes 3B (4-5)*

Display the lines of text in `clinton.lines` that contain “Bill” without a following punctuation mark, or “Trump” with a following punctuation mark.

```
grep("Bill[^[:punct:]]|Trump[[:punct:]]", clinton.lines, value = TRUE)
```

```
## [1] "We heard the man from Hope, Bill Clinton, and the man of hope, Barack
Obama. America is stronger because of President Obama's leadership. And I'm
better because of his friendship."
## [2] "Well, we heard Donald Trump's answer last week at his convention. He
wants to divide us from the rest of the world and from each other. He's betting
```

that the perils of today's world will blind us to its unlimited promise. He's taken the Republican Party a long way, from morning in America to midnight in America."

## [3] "Yes, those were actually Donald Trump's words in Cleveland. And they should set off alarm bells for all of us. Really? I alone can fix it? Isn't he forgetting troops on the front lines, police officers and firefighters who run toward danger, doctors and nurses who care for us, teachers who change lives, entrepreneurs who see possibilities in every problem, mothers who lost children to violence and are building a movement to keep other kids safe? He's forgetting every last one of us."

## [4] "But then I also imagine people are thinking out there, but Trump, he's a businessman, he must know something about the economy."

## [5] "Now, Donald Trump, Donald Trump says, and this is a quote, \"I know more about ISIS than the generals do.\" No, Donald, you don't."

## [6] "And we will stand up against mean and divisive rhetoric wherever it comes from. You know, for the past year many people made the mistake of laughing off Donald Trump's comments, excusing him as an entertainer just putting on a show."

## [7] "Now, at first, at first, I admit, I couldn't he meant it either. It was just too hard to fathom that someone who wants to lead our nation could say those things, could be like that. But here's the sad truth: There is no other Donald Trump, this is it."

## [8] "So enough with the bigotry and the bombast. Donald Trump's not offering real change, he's offering empty promises. And what are we offering? A bold agenda to improve the lives of people across our country to keep you safe, to get you good jobs, to get your kids the opportunities they deserve."

## Question 16

(4 points)

Notes 3B (7-9)

Display the lines of text in `clinton.lines` that contain numbers with at least three digits.

```
grep("[0-9]{3,}", clinton.lines, value = TRUE)
```

## [1] "My friends, we've come to Philadelphia, the birthplace of our nation, because what happened in this city 240 years ago still has something to teach us today. We all know the story, but we usually focus on how it turned out and not enough on how close that story came to never being written at all."

## [2] "I have to tell you, as your secretary of state I went to 112 countries. When people hear those words, they hear America!"

## [3] "240 years later, we still put our faith in each other. Look at what happened in Dallas after the assassinations of five brave police officers. Police Chief David Brown asked the community to support his force, maybe even join them. And you know how the community responded? Nearly 500 people applied in just 12 days."

## [4] "So let's keep going. Let's keep going until every one of the 161 million women and girls across America has the opportunity she deserves to have!"

## [5] "In my first 100 days, we will work with both parties to pass the biggest investment in new, good-paying jobs since World War II. Jobs in manufacturing, clean energy, technology and innovation, small business and infrastructure. If we invest in infrastructure now, we'll not only create jobs today, but lay the foundation for the jobs of the future. And we will also

transform the way we prepare our young people for those jobs."

## Question 17

(4 points)

Notes 3B (4,7)

Display the lines of text in `clinton.lines` that contain a number, then "%" (a percent symbol).

```
grep("[0-9]+%", clinton.lines, value = TRUE)
```

```
## [1] "This is not because we resent success. But when more than 90% of the  
gains have gone to the top 1%, that's where the money is. And we are going to  
follow the money."
```

## Question 18

(4 points)

Notes 3B (6)

Display the lines of text in `clinton.lines` that contain "!", an exclamation mark.

```
grep("!", clinton.lines, value = TRUE)
```

```
## [1] "Thank you all very, very much! Thank you for that amazing welcome!"  
## [2] "To all of you whose hard work brought us here tonight and to those of  
you who joined this campaign this week, thank you. What a remarkable week it's  
been!"  
## [3] "Our country needs your ideas, energy and passion. That is the only way  
we can turn our progressive platform into real change for America. We wrote it  
together, now let's go out and make it happen together!"  
## [4] "He wants us to fear the future and fear each other. Well, you know, a  
great Democratic President Franklin Delano Roosevelt came up with the perfect  
rebuke to Trump more than 80 years ago during a much more perilous time: The  
only thing we have to fear is fear itself!"  
## [5] "I have to tell you, as your secretary of state I went to 112 countries.  
When people hear those words, they hear America!"  
## [6] "Americans don't say \"I alone can fix it.\" We say \"we'll fix it  
together!\""   
## [7] "That's the country we're fighting for. That's the future we're working  
toward. And so, my friends, it is with humility, determination and boundless  
confidence in America's promise that I accept your nomination for president of  
the United States!"  
## [8] "And you heard from Republicans and independents who are supporting our  
campaign. Well, I will be a president for Democrats, Republicans, independents,  
for the struggling, the striving, the successful, for all those who vote for me  
and for those who don't. For all Americans together!"  
## [9] "Tonight we've reached a milestone in our nation's march toward a more  
perfect union. The first time that a major party has nominated a woman for  
president!"  
## [10] "After all, when there are no ceilings, the sky's the limit!"  
## [11] "So let's keep going. Let's keep going until every one of the 161  
million women and girls across America has the opportunity she deserves to  
have!"  
## [12] "And if necessary, we will pass a constitutional amendment to overturn  
Citizens United!"
```



```
## [13] "And I believe in science!"
## [14] "If you believe that companies should share profits, not pad executive
bonuses, join us!"
## [15] "If you believe the minimum wage should be a living wage and no one
working full time should have to raise their children in poverty, join us!"
## [16] "If you believe that every man, woman and child in America has the
right to affordable health care, join us!"
## [17] "If you believe that we should say no to unfair trade deals, that we
should stand up to China, that we should support our steelworkers and
autoworkers and home-grown manufacturers, then join us!"
## [18] "If you believe we should expand Social Security and protect a woman's
right to make her own health care decisions, then join us!"
## [19] "And yes, yes, if you believe that your working mother, wife, sister or
daughter deserves equal pay, join us!"
## [20] "And we will help you balance family and work. And you know what? If
fighting for affordable child care and paid family leave is playing the woman
card, then deal me in!"
## [21] "Now, remember what the president said last night: Don't boo; vote!"
## [22] "A man you can bait with a tweet is not a man we can trust with nuclear
weapons!"
## [23] "And that starts with listening, listening to each other, trying as
best we can to walk in each other's shoes. So let's put ourselves in the shoes
of young black and Latino men and women who face the effects of systemic racism
and are made to feel like their lives are disposable!"
## [24] "And in the end, it comes down to what Donald Trump doesn't get:
America is great because America is good!"
## [25] "Yes, the world is watching what we do. Yes, America's destiny is ours
to choose. So let's be stronger together, my fellow Americans!"
## [26] "Let's look to the future with courage and confidence. Let's build a
better tomorrow for our beloved children and our beloved country. And when we
do, America will be greater than ever!"
## [27] "Thank you, and may God bless you and the United States of America!"
```

## Question 19

(4 points)

Notes 3B (4,6-7)

Display the lines of text in `clinton.lines` that contain a capitalized word followed by a period.

```
grep("[A-Z][a-z]*\\.", clinton.lines, value = TRUE)
```

```
## [1] "We heard the man from Hope, Bill Clinton, and the man of hope, Barack
Obama. America is stronger because of President Obama's leadership. And I'm
better because of his friendship."
## [2] "We heard from our terrific vice president, the one and only Joe Biden."
## [3] "And I want to thank Bernie Sanders."
## [4] "Our country needs your ideas, energy and passion. That is the only way
we can turn our progressive platform into real change for America. We wrote it
together, now let's go out and make it happen together!"
## [5] "Well, we heard Donald Trump's answer last week at his convention. He
wants to divide us from the rest of the world and from each other. He's betting
that the perils of today's world will blind us to its unlimited promise. He's
taken the Republican Party a long way, from morning in America to midnight in
America."
```

## [6] "Yet, we know there is a lot to do. Too many people haven't had a pay raise since the crash. There's too much inequality, too little social mobility, too much paralysis in Washington."

## [7] "Yes, those were actually Donald Trump's words in Cleveland. And they should set off alarm bells for all of us. Really? I alone can fix it? Isn't he forgetting troops on the front lines, police officers and firefighters who run toward danger, doctors and nurses who care for us, teachers who change lives, entrepreneurs who see possibilities in every problem, mothers who lost children to violence and are building a movement to keep other kids safe? He's forgetting every last one of us."

## [8] "Twenty years ago I wrote a book called \"It Takes a Village.\" And a lot of people looked at the title and asked, what the heck do you mean by that? This is what I mean. None of us can raise a family, build a business, heal a community or lift a country totally alone."

## [9] "Because he believed that if he gave everything he had, his children would have a better life than he did. And he was right. My dad, Hugh, made it to college, he played football at Penn State and enlisted in the Navy after Pearl Harbor. When the war was over, he started his own small business printing fabric for draperies. I remember watching him stand for hours over silkscreens. He wanted to give my brothers and me opportunities he never had, and he did."

## [10] "After the four days of this convention, you've seen some of the people who have inspired me, people who let me into their lives and became a part of mine, people like Ryan Moore and Lauren Manning. They told their stories Tuesday night."

## [11] "Lauren Manning, who stood here with such grace and power, was gravely injured on 9/11. It was the thought of her and Debbie St. John who you saw in the movie and John Dolan and Joe Sweeney and all the victims and survivors that kept me working as hard as I could in the Senate on behalf of 9/11 families and our first responders who got sick from their time at ground zero."

## [12] "And in this campaign, I've met many more people who motivate me to keep fighting for change. And with your help, I will carry all of your voices and stories with me to the White House."

## [13] "My primary mission as president will be to create more opportunity and more good jobs with rising wages right here in the United States."

## [14] "In my first 100 days, we will work with both parties to pass the biggest investment in new, good-paying jobs since World War II. Jobs in manufacturing, clean energy, technology and innovation, small business and infrastructure. If we invest in infrastructure now, we'll not only create jobs today, but lay the foundation for the jobs of the future. And we will also transform the way we prepare our young people for those jobs."

## [15] "You know, anyone-anyone-reading the news can see the threats and turbulence we face, from Baghdad to Kabul to Nice and Paris and Brussels, from San Bernardino to Orlando. We're dealing with determined enemies that must be defeated."

## [16] "And I'm proud to stand by our allies in NATO against any threat they face, including from Russia."

## [17] "I've laid out my strategy for defeating ISIS. We will strike their sanctuaries from the air and support local forces taking them out on the ground. We will surge our intelligence so we detect and prevent attacks before they happen. We will disrupt their efforts online to reach and radicalize young people in our country."

## [18] "Well, I've had the privilege to work closely with our troops and our veterans for many years, including as a senator on the Armed Services Committee, and I know how wrong he is. Our military is a national treasure. We

entrust our commander in chief to make the hardest decisions our nation faces, decisions about war and peace, life and death. A president should respect the men and women who risk their lives to serve our country...including Captain Khan and the sons of Tim Kaine and Mike Pence, both Marines."

## [19] "I can't put it any better than Jackie Kennedy did after the Cuban Missile Crisis. She said that what worried President Kennedy during that very dangerous time was that a war might be started not by big men with self-control and restraint, but by little men, the ones moved by fear and pride."

## [20] "I'm not here to repeal the Second Amendment. I'm not here to take away your guns. I just don't want you to be shot by someone who shouldn't have a gun in the first place."

## [21] "That is the story of America. And we begin a new chapter tonight."

---

Below, we read in lines of data from the Advanced National Seismic System (ANSS), on earthquakes of magnitude 6+, between 2002 and 2017. We display the first 15 lines. (You don't have to do anything yet.)

```
anss.lines = readLines("http://www.stat.cmu.edu/~pfreeman/anss.htm")
head(anss.lines,15)

## [1] "<!-- saved from
url=(0047)http://www.ncedc.org/cgi-bin/catalog-search2.pl -->"
## [2] "<html><head><meta http-equiv=\"Content-Type\" content=\"text/html;
charset=windows-1252\"><title>NCEDC_Search_Results</title></head><body>Your
search parameters are:<ul>"
## [3] "<li>catalog=ANSS"
## [4] "</li><li>minimum_magnitude=6.0"
## [5] "</li><li>maximum_magnitude=10"
## [6] "</li><li>event_type=E"
## [7] "</li></ul>"
## [8] "<pre>Date Time Lat Lon Depth Mag Magt Nst Gap Clo RMS SRC Event ID"
## [9]
"-----"
## [10] "2002/01/01 10:39:06.82 -55.2140 -129.0000 10.00 6.00 Mw 78 1.07 NEI
200201014017"
## [11] "2002/01/01 11:29:22.73 6.3030 125.6500 138.10 6.30 Mw 236 0.90 NEI
200201014018"
## [12] "2002/01/02 14:50:33.49 -17.9830 178.7440 665.80 6.20 Mw 215 1.08 NEI
200201024034"
## [13] "2002/01/02 17:22:48.76 -17.6000 167.8560 21.00 7.20 Mw 427 0.90 NEI
200201024041"
## [14] "2002/01/03 07:05:27.67 36.0880 70.6870 129.30 6.20 Mw 431 0.87 NEI
200201034024"
## [15] "2002/01/03 10:17:36.30 -17.6640 168.0040 10.00 6.60 Mw 386 1.14 NEI
200201034040"
```

---

## Question 20

(4 points)

Notes 3B (4,7-9)

This looks like webpage code mixed in with earthquake data. We don't care about the first nine lines, and it looks as if the data we want starts on line 10. Importantly, every line of data contains a date, of the form

YYYY/MM/DD, as in “2002/01/01”. Design a regular expression, call it `date.pattern`, to match to these dates. (Hint: use quantifiers to make `date.pattern` concise.). Use this and `grep()` to retrieve the lines of text containing earthquake data. Call the result `date.lines`. How many lines of data are there? Show the first 2 and the last 2 lines (via the use of `head()` and `tail()`). When was the last earthquake recorded and what was its magnitude?

```
date.pattern = "20[0-9]{2}/[0-9]{2}/[0-9]{2}"
date.lines = grep(date.pattern, anss.lines, value = TRUE)
length(date.lines)
```

```
## [1] 2426
```

```
head(date.lines, 2)
```

```
## [1] "2002/01/01 10:39:06.82 -55.2140 -129.0000 10.00 6.00 Mw 78 1.07 NEI
200201014017"
```

```
## [2] "2002/01/01 11:29:22.73 6.3030 125.6500 138.10 6.30 Mw 236 0.90 NEI
200201014018"
```

```
tail(date.lines, 2)
```

```
## [1] "2017/02/21 14:09:04.41 -19.2839 -63.8990 597.91 6.50 Mw 14 5 0.97 us
201702212014"
```

```
## [2] "2017/02/24 17:28:44.87 -23.2448 -178.8345 414.80 6.90 Mw 19 6 0.96 us
201702242019"
```

There are 2426 lines of data. The last earthquake recorded was in 2017/02/24.  
The magnitude was 6.90.

## Question 21

(4 points)

Notes 3B (4,6-9)

Look back at the lines of earthquake data printed above. The columns for **Lat** and **Lon** give the latitude and longitude, respectively, of the earthquake. Importantly, this takes the form X.XXXX, XX.XXXX, XXX.XXXX, or any of these forms with a leading minus sign, where X is a number. Design a regular expression to match these entries, call it `geo.pattern`. Test it out on the trial string vector below, with `grep()`, and make sure that you match all the strings. (Hint: build the regular expression, “left to right”, following this logic: an optional minus sign, 1 to 3 digits, a period, then exactly 4 digits.)

```
trial.str.vec = c("-55.2140", "-129.0000", "6.3030", "125.6500", "-17.9830")
geo.pattern = "-?[0-9]{1,3}.[0-9]{4}"
grep(geo.pattern, trial.str.vec, value = TRUE)
```

```
## [1] "-55.2140" "-129.0000" "6.3030" "125.6500" "-17.9830"
```

## Question 22

(4 points)

Notes 3B (4,6-9) and Notes 3C (2-3)

Design a regular expression `geo.pattern.sp` that captures not only the latitude/longitude pattern (like `geo.pattern`), but additionally, any number of leading spaces (1 or more). Test it out on the trial string vector below, with `regexpr()` and `regmatches()`. Make sure that you match all the strings, and in each case the extracted text is the entire string.

```
trial.str.vec.sp = c(" -55.2140", " -129.0000", " 6.3030", " 125.6500", " -17.9830")
geo.pattern.sp = " +-[0-9]{1,3}.[0-9]{4}"
reg_exp = regexpr(geo.pattern.sp, trial.str.vec.sp, useBytes = TRUE)
regmatches(trial.str.vec.sp, reg_exp)
```

```
## [1] " -55.2140" " -129.0000" " 6.3030" " 125.6500" " -17.9830"
```

## Question 23

(4 points)

Notes 3B (4,6-9) and Notes 3C (2-3)

Finally, design a regular expression `geo.pattern.pair` that captures a latitude pattern, then any number of spaces (1 or more), then a longitude pattern. Really, this is just the concatenation of the regexes you already designed, `geo.pattern` and `geo.pattern.sp`. Use `geo.pattern.pair`, with `regexpr()` and `regmatches()`, in order to extract the latitude/longitude pairs from each line of earthquake data in `date.lines`. Call the result `lat.lon.pairs`, and display the first 3 entries, checking visually that it matches the results printed at the top of this lab.

```
geo.pattern.pair = paste(geo.pattern, geo.pattern.sp, sep = " ")
reg_exp = regexpr(geo.pattern.pair, date.lines, useBytes = TRUE)
lat.lon.pairs = regmatches(date.lines, reg_exp)
head(lat.lon.pairs, 3)
```

```
## [1] "-55.2140 -129.0000" "6.3030 125.6500" "-17.9830 178.7440"
```

## Question 24

(4 points)

Notes 3B (7-8,11)

Apply `strsplit()` to `lat.lon.pairs[1:2]` with `split` set equal to a regex that captures any number of spaces (1 or more), and show the results. Check that the two elements of the return list have the same number of strings. You'll find that the results have an extra, empty string. Clean this up by doing the following: (a) apply `unlist()` to `lat.lon.pairs[1:2]` and save the result to a variable; (b) identify the indices of all elements of your new vector variable for which the value is the empty string; (c) remove those elements; and (d) put the clean-up vector into a two-row matrix. Note: (b) and (c) can be effectively combined into one expression. Also note that since matrices get populated by filling the first column, then the second column, etc., you'll have to display the transpose of the matrix you define for step (d).

```
res = strsplit(lat.lon.pairs[1:2], split = " +")
res
```

```
## [[1]]
## [1] "-55.2140" "-129.0000"
##
## [[2]]
## [1] "6.3030" "125.6500"
```

```
flattened = unlist(lat.lon.pairs[1:2])
t(matrix(flattened, ncol = 2))
```

```
##      [,1]
## [1,] "-55.2140 -129.0000"
## [2,] "6.3030 125.6500"
```

## Question 25

(4 points)

Notes 3B (3-9)

Design a regex to match all patterns of the following form: any letter (1 or more), then a punctuation mark, then “ve” or “ll” or “t”, then a space or a punctuation mark. Call it `my.pattern`. Check, with `grep()`, that it matches all the strings below.

```
trial.str.vec = c("we've.", "I'll ", "don't!")
my.pattern = "[[:alpha:]]+[[:punct:]](ve|ll|t)( |[:punct:]))"
grep(my.pattern, trial.str.vec, value = TRUE)
```

```
## [1] "we've." "I'll "  "don't!"
```