

# Project 20

Name: Elizabeth McGuckin

Using RStudio, we start to learn how to extract data from the web.

Use the data from the Billboard Hot 100 for question 1. Please use the data from the week you were born. For instance, if I solve question 1, I would use the data located here: <https://www.billboard.com/charts/hot-100/1976-10-13>

On the Hot 100 chart, from the day of your birth:

**1a.** Extract the titles of the songs ranked #2 through #100.

```
library(XML)
library(RCurl)

## Loading required package: bitops

xpathSApply(htmlParse(getURL("https://www.billboard.com/charts/hot-100/1998-06-18")),
             "//*[@div[@class='chart-list-item__title']]", xmlValue)

## [1] "\n\nYou're Still The One\n\n"
## [2] "\n\nToo Close\n\n"
## [3] "\n\nMy All\n\n"
## [4] "\n\nI Get Lonely\n\n"
## [5] "\n\nEverybody [Backstreet's Back]\n\n"
## [6] "\n\nThey Don't Know\n\n"
## [7] "\n\nAll My Life\n\n"
## [8] "\n\nAdia\n\n"
## [9] "\n\nTruly Madly Deeply\n\n"
## [10] "\n\nThe Arms Of The One Who Loves You\n\n"
## [11] "\n\nSex And Candy\n\n"
## [12] "\n\nMy Way\n\n"
## [13] "\n\nBody Bumpin' Yippie-Yi-Yo\n\n"
## [14] "\n\nSay It\n\n"
## [15] "\n\nIt's All About Me\n\n"
## [16] "\n\nLet's Ride\n\n"
## [17] "\n\nI Got The Hook Up!\n\n"
## [18] "\n\nI Want You Back\n\n"
## [19] "\n\nTurn It Up [Remix]/Fire It Up\n\n"
## [20] "\n\nLooking Through Your Eyes\n\n"
## [21] "\n\nFrozen\n\n"
## [22] "\n\nGettin' Jiggy Wit It\n\n"
## [23] "\n\nMake Em' Say Uhh!\n\n"
## [24] "\n\nNo, No, No\n\n"
## [25] "\n\nHow Do I Live\n\n"
## [26] "\n\nHeaven's What I Feel\n\n"
## [27] "\n\nHow's It Going To Be\n\n"
## [28] "\n\nTogether Again\n\n"
## [29] "\n\nShorty (You Keep Playin' With My Mind)\n\n"
## [30] "\n\nThis Kiss\n\n"
## [31] "\n\nRaise The Roof\n\n"
## [32] "\n\nI Don't Want To Wait\n\n"
## [33] "\n\nMoney, Power & Respect\n\n"
## [34] "\n\nVictory\n\n"
```

## [35] "\n\nStop\n\n"  
 ## [36] "\n\nWhatcha Gone Do?\n\n"  
 ## [37] "\n\nCheers 2 U\n\n"  
 ## [38] "\n\n Still Not A Player\n\n"  
 ## [39] "\n\nNice & Slow\n\n"  
 ## [40] "\n\nWhat You Want\n\n"  
 ## [41] "\n\nYou Make Me Wanna...\n\n"  
 ## [42] "\n\nWho Am I\n\n"  
 ## [43] "\n\nMy Heart Will Go On\n\n"  
 ## [44] "\n\nCandle In The Wind 1997/Something About The Way You Look Tonight\n\n"  
 ## [45] "\n\nImagination\n\n"  
 ## [46] "\n\nGet At Me Dog\n\n"  
 ## [47] "\n\nYou Won't Forget Me\n\n"  
 ## [48] "\n\nDeja Vu [Uptown Baby]\n\n"  
 ## [49] "\n\nGone Till November\n\n"  
 ## [50] "\n\nRomeo And Juliet\n\n"  
 ## [51] "\n\nWhen The Lights Go Out\n\n"  
 ## [52] "\n\nWishlist\n\n"  
 ## [53] "\n\nBitter Sweet Symphony\n\n"  
 ## [54] "\n\nPush It\n\n"  
 ## [55] "\n\n Second Round K.O.\n\n"  
 ## [56] "\n\nClock Strikes\n\n"  
 ## [57] "\n\nOne Heart At A Time\n\n"  
 ## [58] "\n\nRecover Your Soul\n\n"  
 ## [59] "\n\nThere's Your Trouble\n\n"  
 ## [60] "\n\nI'm Alright\n\n"  
 ## [61] "\n\nNinety Nine [Flash The Message]\n\n"  
 ## [62] "\n\nThe Mummers' Dance\n\n"  
 ## [63] "\n\nLove You Down\n\n"  
 ## [64] "\n\nThis Is How We Party\n\n"  
 ## [65] "\n\nDo Your Thing\n\n"  
 ## [66] "\n\nI Honestly Love You (David Foster Vers.)\n\n"  
 ## [67] "\n\nThe Party Continues\n\n"  
 ## [68] "\n\nA Rose Is Still A Rose\n\n"  
 ## [69] "\n\nDo For Love\n\n"  
 ## [70] "\n\nDing-A-Ling\n\n"  
 ## [71] "\n\nI Just Want To Dance With You\n\n"  
 ## [72] "\n\nGitty Up\n\n"  
 ## [73] "\n\nJust Be Straight With Me\n\n"  
 ## [74] "\n\nNow That I Found You\n\n"  
 ## [75] "\n\nI Do [Cherish You]\n\n"  
 ## [76] "\n\nI'm From The Country\n\n"  
 ## [77] "\n\nMy Heart Will Go On\n\n"  
 ## [78] "\n\nThe Unforgiven II\n\n"  
 ## [79] "\n\nSin So Well\n\n"  
 ## [80] "\n\nA Man Holdin' On (To A Woman Lettin' Go)\n\n"  
 ## [81] "\n\nHooked On A Feeling (Ooga-Chaka)\n\n"  
 ## [82] "\n\nParty Ain't A Party\n\n"  
 ## [83] "\n\nFreak Out\n\n"  
 ## [84] "\n\nKnow What You Mean\n\n"  
 ## [85] "\n\nOne Of These Days\n\n"  
 ## [86] "\n\nToo Much\n\n"  
 ## [87] "\n\nWhatuon\n\n"  
 ## [88] "\n\nBaby Be There\n\n"

```
## [89] "\n\nStill Po' Pimpin'\n\n"
## [90] "\n\nOut Of My Bones\n\n"
## [91] "\n\nYou Only Have To Say You Love Me\n\n"
## [92] "\n\nI Wanna Get Next To You\n\n"
## [93] "\n\nBlue On Black\n\n"
## [94] "\n\nThere You Are\n\n"
## [95] "\n\nAre You Jimmy Ray?\n\n"
## [96] "\n\nRain\n\n"
## [97] "\n\nLight In Your Eyes\n\n"
## [98] "\n\nIt's Your Love\n\n"
## [99] "\n\nThat's Why I'm Here\n\n "
```

Explain solution: Changed the code to include my birthday so that it found the top 2-99 songs on my birthday.

1b. Extract the artists for those 99 songs.

```
xpathSApply(htmlParse(getURL("https://www.billboard.com/charts/hot-100/1998-06-18")),
             "//*[div[@class='chart-list-item__artist']]", xmlValue)
```

```
## [1] "\n\nShania Twain\n\n"
## [2] "\n\nNext\n\n"
## [3] "\n\nMariah Carey\n\n"
## [4] "\n\nJanet Featuring BLACKstreet\n\n"
## [5] "\n\nBackstreet Boys\n\n"
## [6] "\n\nJon B\n\n"
## [7] "\n\nK-Ci & JoJo\n\n"
## [8] "\n\nSarah McLachlan\n\n"
## [9] "\n\nSavage Garden\n\n"
## [10] "\n\nXscape\n\n"
## [11] "\n\nMarcy Playground\n\n"
## [12] "\n\nUsher\n\n"
## [13] "\n\nPublic Announcement\n\n"
## [14] "\n\nVoices Of Theory\n\n"
## [15] "\n\nMya & Sisqo\n\n"
## [16] "\n\nMontell Jordan Feat. Master P & Silkk The Shocker\n\n"
## [17] "\n\nMaster P Featuring Sons Of Funk\n\n"
## [18] "\n\n'N Sync\n\n"
## [19] "\n\nBusta Rhymes\n\n "
## [20] "\n\nLeAnn Rimes\n\n"
## [21] "\n\nMadonna\n\n"
## [22] "\n\nWill Smith\n\n"
## [23] "\n\nMaster P Feat. Fiend, Silkk The Shocker, Mia X & Mystikal\n\n"
## [24] "\n\nDestiny's Child\n\n"
## [25] "\n\nLeAnn Rimes\n\n"
## [26] "\n\nGloria Estefan\n\n"
## [27] "\n\nThird Eye Blind\n\n"
## [28] "\n\nJanet\n\n"
## [29] "\n\nImajin Featuring Keith Murray\n\n"
## [30] "\n\nFaith Hill\n\n"
## [31] "\n\nLuke Featuring No Good But So Good\n\n"
## [32] "\n\nPaula Cole\n\n"
## [33] "\n\nThe Lox [Featuring DMX & Lil' Kim]\n\n"
## [34] "\n\nPuff Daddy & The Family Featuring The Notorious B.I.G. & Busta Rh\n\n"
## [35] "\n\nSpice Girls\n\n"
## [36] "\n\nLink\n\n"
```

## [37] "\n\nPlaya\n\n"  
 ## [38] "\n\nBig Punisher Featuring Joe\n"  
 ## [39] "\n\nUsher\n\n"  
 ## [40] "\n\nMase Featuring Total\n"  
 ## [41] "\n\nUsher\n\n"  
 ## [42] "\n\nBeenie Man\n\n"  
 ## [43] "\n\nCeline Dion\n\n"  
 ## [44] "\n\nElton John\n\n"  
 ## [45] "\n\nTamia\n\n"  
 ## [46] "\n\nDMX Featuring Sheek Of The Lox\n"  
 ## [47] "\n\nLa Bouche\n\n"  
 ## [48] "\n\nLord Tariq & Peter Gunz\n"  
 ## [49] "\n\nWyclef Jean\n\n"  
 ## [50] "\n\nSylk-E. Fyne Featuring Chill\n"  
 ## [51] "\n\nFive\n\n"  
 ## [52] "\n\nPearl Jam\n\n"  
 ## [53] "\n\nThe Verve\n\n"  
 ## [54] "\n\nGarbage\n\n"  
 ## [55] "\n\nCanibus\n\n"  
 ## [56] "\n\nTimbaland & Magoo\n\n"  
 ## [57] "\n\nVarious Artists\n\n"  
 ## [58] "\n\nElton John\n\n"  
 ## [59] "\n\nDixie Chicks\n\n"  
 ## [60] "\n\nJo Dee Messina\n\n"  
 ## [61] "\n\nJohn Forte\n\n"  
 ## [62] "\n\nLoreena McKennitt\n\n"  
 ## [63] "\n\nINOJ/LATHUN\n"  
 ## [64] "\n\nS.O.A.P.\n\n "  
 ## [65] "\n\n7 Mile\n\n"  
 ## [66] "\n\nOlivia Newton-John\n\n"  
 ## [67] "\n\nJD Featuring Da Brat\n"  
 ## [68] "\n\nAretha Franklin\n\n"  
 ## [69] "\n\n2Pac Featuring Eric Williams\n"  
 ## [70] "\n\nHi-Town DJs\n\n"  
 ## [71] "\n\nGeorge Strait\n\n"  
 ## [72] "\n\nSalt-N-Pepa\n\n"  
 ## [73] "\n\nSilkk The Shocker Featuring Master P, Destiny's Child, O'Dell, Mo\n"  
 ## [74] "\n\nTerri Clark\n\n"  
 ## [75] "\n\nMark Wills\n\n"  
 ## [76] "\n\nTracy Byrd\n\n"  
 ## [77] "\n\nDeja Vu\n\n"  
 ## [78] "\n\nMetallica\n\n"  
 ## [79] "\n\nRebekah\n\n"  
 ## [80] "\n\nTy Herndon\n\n"  
 ## [81] "\n\nBaby Talk\n\n"  
 ## [82] "\n\nQueen Pen Featuring Teddy Riley, Nutta Butta, Markell & Jesse Wes\n"  
 ## [83] "\n\nNutta Butta Featuring Teddy Riley And Anonymous\n"  
 ## [84] "\n\nSister 7\n\n"  
 ## [85] "\n\nTim McGraw\n\n"  
 ## [86] "\n\nSpice Girls\n\n"  
 ## [87] "\n\nLaTanya Featuring Twista\n"  
 ## [88] "\n\nNu Flavor\n\n"  
 ## [89] "\n\nDo Or Die Featuring Johnny P & Twista\n"  
 ## [90] "\n\nRandy Travis\n\n"

```
## [91] "\n\nHannah Jones\n\n"
## [92] "\n\nChristion\n\n"
## [93] "\n\nThe Kenny Wayne Shepherd Band\n\n"
## [94] "\n\nSam Salter\n\n"
## [95] "\n\nJimmy Ray\n\n"
## [96] "\n\nSWV\n\n"
## [97] "\n\nBlessid Union Of Souls\n\n"
## [98] "\n\nShe Moves\n\n"
## [99] "\n\nKenny Chesney\n\n"
```

Explain solution: In this line of code we found the artists for the top songs on my birthday.

**1c.** Extract the title of the number 1 song for that day.

```
xpathSApply(htmlParse(getURL("https://www.billboard.com/charts/hot-100/1998-06-18")),
             "//*[@div[@class='chart-number-one__title']]", xmlValue)
```

```
## [1] "The Boy Is Mine"
```

Explain solution: In this line of code I changed the class because I was looking for the number one song.

**1d.** Extract the artist for the number 1 song for that day.

```
xpathSApply(htmlParse(getURL("https://www.billboard.com/charts/hot-100/1976-10-13")),
             "//*[@div[@class='chart-number-one__artist']]", xmlValue)
```

```
## [1] "\n\nRick Dees & His Cast Of Idiots\n\n"
```

Explain solution: In this line of code we changed the class to find the number one artist.

**2a.** Extract the city where the National Park property for Catoclin Mountain is located. This data is found at: <https://www.nps.gov/cato/index.htm> or in the file: /depot/statclass/data/parks/cato.htm

```
xpathSApply(htmlParse(getURL("https://www.nps.gov/cato/index.htm")), "//*[@span[@itemprop='addressLocality']]", xmlValue)
```

```
## [1] "Thurmont"
```

Explain solution: By using addressLocality I was able to see the city, cato let me see the cities in Catoclin Mountain.

**2b.** Extract the state where Catoclin Mountain is located.

```
xpathSApply(htmlParse(getURL("https://www.nps.gov/cato/index.htm")), "//*[@span[@itemprop='addressRegion']]", xmlValue)
```

```
## [1] "MD"
```

Explain solution: By using addressRegion I was able to see the state.

**2c.** Extract the zip code where Catoclin Mountain is located.

```
xpathSApply(htmlParse(getURL("https://www.nps.gov/cato/index.htm")), "//*[@span[@itemprop='postalCode']]", xmlValue)
```

```
## [1] "21788      "
```

Explain solution: By using postalCode I was able to see the postal code for this location.

**3a.** Identify three potential websites that you are interested to try to scrape yourself, during the upcoming seminars. Look for websites with data that is (relatively) easy to scrape, for instance: Systematic URL's that are easy to understand; (relative) consistency in how the data is stored; and make sure that the data is embedded in the page, rather than in csv files that are already prepared for download. (We want to actually scrape some data.) 1. <https://www.ssa.gov/oact/babynames/> 2. <http://www.all-my-favourite-flower-names.com/list-of-flower-names.html> 3. [https://digitaldreamdoor.com/pages/music0\\_christmas.html](https://digitaldreamdoor.com/pages/music0_christmas.html) **3b.** For each of the three websites that you identified, give a very brief description of the kind of data that you want to scrape.

1. I want to scrape baby names
2. I want to scrape flower name
3. I want to scrape christmas music