

Lab 4 Submission

Data Wrangler Script and Screenshot: CMSC

rows: 3

| Course_number | Section_Number | Instructor | Open | Waitlist | |
|---------------|----------------|----------------|------|----------|------|
| 1 CMSC100 | 0101 | Charles Kassir | 45 | 0 | M |
| 2 CMSC106 | 0101 | Jianwu Wang | 45 | 5 | TuTh |
| 3 CMSC131 | 0101 | Evan Golub | 31 | 0 | MWF |

| Open | Waitlist | Days | Time | Building | Room_No. |
|------|----------|------|------------------|----------|----------|
| 0 | | M | 4:00pm - 4:50pm | CSI | 2117 |
| 5 | | TuTh | 9:30am - 10:45am | CSI | 2117 |
| 0 | | MWF | 2:00pm - 2:50pm | CSI | 1115 |

rows: 52 prev next

| Course_number | Section_Number | Instructor | Open | Waitlist | |
|---------------|----------------|--------------------------|------|----------|------|
| 1 CMSC100 | 0101 | Charles Kassir | 45 | 0 | M |
| 2 CMSC106 | 0101 | Jianwu Wang | 45 | 5 | TuTh |
| 3 CMSC131 | 0101 | Evan Golub | 31 | 0 | MWF |
| 4 CMSC132 | 0101 | Laurence Herman | 34 | 2 | MWF |
| 5 CMSC216 | 0101 | Nelson Padua-Perez | 28 | 0 | TuTh |
| 6 CMSC250 | 0101 | Clyde Kruskal | 29 | 4 | TuTh |
| 7 CMSC250 | 0101 | Thomas Goldstein | 25 | 0 | TuTh |
| 8 CMSC289 | 0101 | James Reggia | 60 | 36 | TuTh |
| 9 CMSC330 | 0101 | Chau-Wen Tseng | 27 | 0 | TuTh |
| 10 CMSC351 | 0101 | Hamid Mahini | 88 | 2 | MWF |
| 11 CMSC396 | 0101 | Atif Memon , Neil Spring | 25 | 0 | W |
| 12 CMSC411 | 0101 | Michelle Hugue | 45 | 3 | TuTh |
| 13 CMSC412 | 0101 | Neil Spring | 25 | 4 | TuTh |
| 14 CMSC414 | 0101 | A.U. Shankar | 40 | 6 | TuTh |
| 15 CMSC417 | 0101 | Ashok Agrawala | 40 | 0 | TuTh |
| 16 CMSC420 | 0101 | Hanan Samet | 40 | 0 | TuTh |
| 17 CMSC421 | 0101 | Donald Perlis | 50 | 2 | TuTh |
| 18 CMSC423 | 0101 | Hector Corrada Bravo | 48 | 0 | MW |

Data Wrangler Script and Screenshot: World Cup 1

row

data = '12 ([[1938 FIFA Wo

suggestions

Delete row 4

Delete rows where data = '12 ([[1938 FIFA World Cup|1938...

Promote row 4 to header

Fill row 4 with values from above

Fill row 4 with values from below

Fill rows where data = '12 ([[1938 FIFA World Cup|1938... with values from above

Fill rows where data = '12 ([[1938 FIFA World Cup|1938... with values from below

ScriptExport

Split data repeatedly on newline into rows

Delete row 1

Delete rows where data = '12

Delete every 6 rows starting with 6

rows: 122prevnext

data

1|style="background: #fff66f"|{{fb|BRA}}

2|style="background: #fff66f"|5 ([[1958 FIFA World Cup|1958]], [[1962 FIFA World Cup|1962]], [[1970 FIFA World Cup|1970]], [[1994 FIFA World Cup|1994]], [[2002 FIFA World Cup|2002]])

3|2 ([[1950 FIFA World Cup|1950]]|{{#1|*}}), [[1998 FIFA World Cup|1998]])

4|2 ([[1938 FIFA World Cup|1938]], [[1978 FIFA World Cup|1978]])

5|2 ([[1974 FIFA World Cup|1974]], [[2014 FIFA World Cup|2014]]|{{#1|*}})

6|style="background: #fff66f"|{{fb|GER}}|{{#2|*}}

7|style="background: #fff66f"|4 ([[1954 FIFA World Cup|1954]], [[1974 FIFA World Cup|1974]]|{{#1|*}}), [[1990 FIFA World Cup|1990]], [[2014 FIFA World Cup|2014]])

8|4 ([[1966 FIFA World Cup|1966]], [[1982 FIFA World Cup|1982]], [[1986 FIFA World Cup|1986]], [[2002 FIFA World Cup|2002]])

9|4 ([[1934 FIFA World Cup|1934]], [[1970 FIFA World Cup|1970]], [[2006 FIFA World Cup|2006]]|{{#1|*}}), [[2010 FIFA World Cup|2010]])

10|1 ([[1958 FIFA World Cup|1958]])

11|style="background: #fff66f"|{{fb|ITA}}

12|style="background: #fff66f"|4 ([[1934 FIFA World Cup|1934]]|{{#1|*}}), [[1938 FIFA World Cup|1938]], [[1982 FIFA World Cup|1982]], [[2006 FIFA World Cup|2006]])

13|2 ([[1970 FIFA World Cup|1970]], [[1994 FIFA World Cup|1994]])

14|1 ([[1990 FIFA World Cup|1990]]|{{#1|*}})

15|1 ([[1978 FIFA World Cup|1978]])

16|style="background: #fff66f"|{{fb|ARG}}

17|style="background: #fff66f"|2 ([[1978 FIFA World Cup|1978]]|{{#1|*}}), [[1986 FIFA World Cup|1986]])

18|3 ([[1930 FIFA World Cup|1930]], [[1990 FIFA World Cup|1990]], [[2014 FIFA World Cup|2014]])

19|align=center|{{sort dash}}

20|align=center|{{sort dash}}

21|style="background: #fff66f"|{{fb|URU}}

22|style="background: #fff66f"|2 ([[1930 FIFA World Cup|1930]]|{{#1|*}}), [[1950 FIFA World Cup|1950]])

23|align=center|{{sort dash}}

suggestions

rows: 122prevnext

data

1|BRA

2|style="background: #fff66f"|5 ([[1958 FIFA World Cup|1958]], [[1962 FIFA World Cup|1962]], [[1970 FIFA World Cup|1970]], [[1994 FIFA World Cup|1994]], [[2002 FIFA World Cup|2002]])

3|2 ([[1950 FIFA World Cup|1950]]|{{#1|*}}), [[1998 FIFA World Cup|1998]])

4|2 ([[1938 FIFA World Cup|1938]], [[1978 FIFA World Cup|1978]])

5|2 ([[1974 FIFA World Cup|1974]], [[2014 FIFA World Cup|2014]]|{{#1|*}})

6|style="background: #fff66f"|{{fb|GER}}|{{#2|*}}

7|style="background: #fff66f"|4 ([[1954 FIFA World Cup|1954]], [[1974 FIFA World Cup|1974]]|{{#1|*}}), [[1990 FIFA World Cup|1990]], [[2014 FIFA World Cup|2014]])

8|4 ([[1966 FIFA World Cup|1966]], [[1982 FIFA World Cup|1982]], [[1986 FIFA World Cup|1986]], [[2002 FIFA World Cup|2002]])

9|4 ([[1934 FIFA World Cup|1934]], [[1970 FIFA World Cup|1970]], [[2006 FIFA World Cup|2006]]|{{#1|*}}), [[2010 FIFA World Cup|2010]])

10|1 ([[1958 FIFA World Cup|1958]])

11|ITA

12|style="background: #fff66f"|4 ([[1934 FIFA World Cup|1934]]|{{#1|*}}), [[1938 FIFA World Cup|1938]], [[1982 FIFA World Cup|1982]], [[2006 FIFA World Cup|2006]])

13|2 ([[1970 FIFA World Cup|1970]], [[1994 FIFA World Cup|1994]])

14|1 ([[1990 FIFA World Cup|1990]]|{{#1|*}})

15|1 ([[1978 FIFA World Cup|1978]])

16|ARG

17|style="background: #fff66f"|2 ([[1978 FIFA World Cup|1978]]|{{#1|*}}), [[1986 FIFA World Cup|1986]])

18|3 ([[1930 FIFA World Cup|1930]], [[1990 FIFA World Cup|1990]], [[2014 FIFA World Cup|2014]])

19|align=center|{{sort dash}}

20|align=center|{{sort dash}}

21|style="background: #fff66f"|{{fb|URU}}

22|style="background: #fff66f"|2 ([[1930 FIFA World Cup|1930]]|{{#1|*}}), [[1950 FIFA World Cup|1950]])

23|align=center|{{sort dash}}

ScriptExport

* Extract from data before '1 any uppercase word '2'

* Cut from data before 'BRA'

* Cut from data on '}}'

* Cut from data before 'ITA'

* Cut from data before 'ARG'

suggestions

rows: 122prevnext

data

1|BRA

2|([[1958 FIFA World Cup|1958]], [[1962 FIFA World Cup|1962]], [[1970 FIFA World Cup|1970]], [[1994 FIFA World Cup|1994]], [[2002 FIFA World Cup|2002]])

3|2 ([[1950 FIFA World Cup|1950]]|{{#1|*}}), [[1998 FIFA World Cup|1998]])

4|2 ([[1938 FIFA World Cup|1938]], [[1978 FIFA World Cup|1978]])

5|2 ([[1974 FIFA World Cup|1974]], [[2014 FIFA World Cup|2014]]|{{#1|*}})

6|style="background: #fff66f"|{{fb|GER}}|{{#2|*}}

7|([[1954 FIFA World Cup|1954]], [[1974 FIFA World Cup|1974]]|{{#1|*}}), [[1990 FIFA World Cup|1990]], [[2014 FIFA World Cup|2014]])

8|4 ([[1966 FIFA World Cup|1966]], [[1982 FIFA World Cup|1982]], [[1986 FIFA World Cup|1986]], [[2002 FIFA World Cup|2002]])

9|4 ([[1934 FIFA World Cup|1934]], [[1970 FIFA World Cup|1970]], [[2006 FIFA World Cup|2006]]|{{#1|*}}), [[2010 FIFA World Cup|2010]])

10|1 ([[1958 FIFA World Cup|1958]])

11|ITA

12|([[1934 FIFA World Cup|1934]]|{{#1|*}}), [[1938 FIFA World Cup|1938]], [[1982 FIFA World Cup|1982]], [[2006 FIFA World Cup|2006]])

13|2 ([[1970 FIFA World Cup|1970]], [[1994 FIFA World Cup|1994]])

14|1 ([[1990 FIFA World Cup|1990]]|{{#1|*}})

15|1 ([[1978 FIFA World Cup|1978]])

16|ARG

17|([[1978 FIFA World Cup|1978]]|{{#1|*}}), [[1986 FIFA World Cup|1986]])

18|3 ([[1930 FIFA World Cup|1930]], [[1990 FIFA World Cup|1990]], [[2014 FIFA World Cup|2014]])

19|align=center|{{sort dash}}

20|align=center|{{sort dash}}

21|style="background: #fff66f"|{{fb|URU}}

22|([[1930 FIFA World Cup|1930]]|{{#1|*}}), [[1950 FIFA World Cup|1950]])

23|align=center|{{sort dash}}

ScriptExport

> Cut from data on '}}'

> Cut from data before 'ITA'

> Cut from data before 'ARG'

> Cut from data on '1 any lowercase word =' # any lowercase word any number any lowercase word '\ ' any number'

| Suggestions | rows: 122prevnext |
|-------------|---|
| | <div>data</div> |
| | 1BRA |
| | 21958 1962 1970 1994 2002 |
| | 31950 FIFA World Cup[1950]][[#1*]], [[1998 FIFA World Cup[1998]]] |
| | 41938 FIFA World Cup[1938]], [[1978 FIFA World Cup[1978]]] |
| | 51974 FIFA World Cup[1974]], [[2014 FIFA World Cup[2014]][[#1*]]] |
| | 6⇑GER[[#2*]] |
| | 71974 FIFA World Cup[1974]][[#1*]], [[1990 FIFA World Cup[1990]], [[2014 FIFA World Cup[2014]]] |
| | 84 1966 FIFA World Cup[1966]], [[1982 FIFA World Cup[1982]], [[1986 FIFA World Cup[1986]], [[2002 |
| | 94 1934 1970 FIFA World Cup[1970]], [[2006 FIFA World Cup[2006]][[#1*]], [[2010 FIFA World |
| | Cup[2010]]] |
| | 101 1958 FIFA World Cup[1958]]] |
| | 11ITA |
| | 121938 FIFA World Cup[1938]], [[1982 FIFA World Cup[1982]], [[2006 FIFA World Cup[2006]]] |
| | 132 1970 1994 FIFA World Cup[1994]]] |
| | 141 1990 FIFA World Cup[1990]][[#1*]]] |
| | 151 1978 FIFA World Cup[1978]]] |
| | 16ARG |
| | 171978 1986 FIFA World Cup[1986]]] |
| | 183 1930 FIFA World Cup[1930]], [[1990 FIFA World Cup[1990]], [[2014 FIFA World Cup[2014]]] |
| | 19[align=center][[sort dash |
| | 20[align=center][[sort dash |
| | 21⇑URU |
| | 221950 FIFA World Cup[1950]]] |
| | 23[align=center][[sort dash |
| | 24[align=center][[sort dash |
| | 253 1954 1970 FIFA World Cup[1970]], [[2010 FIFA World Cup[2010]]] |

ScriptExport

Cut from data between '1994 ' and '2002'

Cut from data after '2002'

Cut from data before '1950'

Cut from data before '1938 '

Cut from data before '1974'

| Suggestions | rows: 122prevnext |
|-------------|---|
| | <div>data</div> |
| | 1BRA |
| | 21958 1962 1970 1994 2002 |
| | 31950 1998 |
| | 41938 1978 |
| | 51974 2014 |
| | 6GER |
| | 71974 2014 |
| | 84 1966 FIFA World Cup[1966]], [[1982 FIFA World Cup[1982]], [[1986 FIFA World Cup[1986]], [[2002 |
| | 94 1934 1970 FIFA World Cup[1970]], [[2006 FIFA World Cup[2006]][[#1*]], [[2010 FIFA World |
| | Cup[2010]]] |
| | 101 1958 FIFA World Cup[1958]]] |
| | 11ITA |
| | 121938 FIFA World Cup[1938]], [[1982 FIFA World Cup[1982]], [[2006 FIFA World Cup[2006]]] |
| | 132 1970 1994 FIFA World Cup[1994]]] |
| | 141 1990 FIFA World Cup[1990]][[#1*]]] |
| | 151 1978 |
| | 16ARG |
| | 171978 |
| | 183 2014 |
| | 19[align=center][[sort dash |
| | 20[align=center][[sort dash |
| | 21⇑URU |
| | 221950 FIFA World Cup[1950]]] |
| | 23[align=center][[sort dash |
| | 24[align=center][[sort dash |
| | 253 1954 1970 FIFA World Cup[1970]], [[2010 FIFA World Cup[2010]]] |

ScriptExport

Cut from data between ' and '1978

Cut from data after '1978'

Cut from data between ' ' and '2014'

Cut from data after '2014'

Cut from data before 'GER'

Cut from data after 'GER'

Data Wrangler Script and Screenshot: World Cup 2

columnmeasure

split2split1

uggestions

rows: 31prevnext

split#split1#split2#

1BRA19621

2BRA19701

3BRA19941

4BRA20021

5BRA19581

6BRA19982

7BRA19502

8BRA19383

9BRA19783

10BRA19744

rows: 3

split_1_2_3_4

1BRA1958195019782014

2GER2014200220061958

3ITA2006199419901978

criptExport

Split data repeatedly on newline into rows

Split data repeatedly on ','

columnkeys

_1962, _1970, _1994, _20020

Suggestions

rows: 3

Countries#_1962#_1970#_1994#_2002#

1BRA11111

2GER324

3ITA2

Fold _1962, _1970, _1994, _2002... using header as a key

Fold _1962, _1970, _1994, _2002... using 1 as a key

Fold _1962, _1970, _1994, _2002... using 1, 2 as keys

columnkeys

_1962, _1970, _1994, _2002

0

Suggestions

Fold **_1962, _1970, _1994, _2002...** using **header** as a key

Fold **_1962, _1970, _1994, _2002...** using **1** as a key

Fold **_1962, _1970, _1994, _2002...** using **1, 2** as keys

ScriptExport

- Split data repeatedly on **newline** into **rows**
- Split data repeatedly on **','**
- Unfold **split1** on **split2**
- Set **split** name to **Countries**

rows: 3

| | Countries | _1962 | _1970 | _1994 | _2002 |
|---|-----------|-------|-------|-------|-------|
| 1 | BRA | 1 | 1 | 1 | 1 |
| 2 | GER | | 3 | | 4 |
| 3 | ITA | | 2 | 2 | |

rows: 54prevnext

| | Countries | fold | value |
|----|-----------|-------|-------|
| 1 | BRA | _1962 | 1 |
| 2 | BRA | _1970 | 1 |
| 3 | BRA | _1994 | 1 |
| 4 | BRA | _2002 | 1 |
| 5 | BRA | _1958 | 1 |
| 6 | BRA | _1998 | 2 |
| 7 | BRA | _1950 | 2 |
| 8 | BRA | _1938 | 3 |
| 9 | BRA | _1978 | 3 |
| 10 | BRA | _1974 | 4 |
| 11 | BRA | _2014 | 4 |
| 12 | BRA | _1954 | |
| 13 | BRA | _1990 | |
| 14 | BRA | _1966 | |
| 15 | BRA | _1982 | |

UNIX Tools Command: CMSC

```
cat cmisc.txt | awk -F'[ ,\t]' 'BEGIN {print "Course No.\tSection No.\tInstructor\tSeats\tOpen\tWaitlist\tDays  
Time\t\t\tBldg.\tRoom No." /^CMSC[0-9][0-9][0-9]$/ {print $1;} /^0[0-9][0-9][0-9]$/ {print "\t\t\t", $0;} /^[A-Za-z]+([A-  
Za-z])+$/ {print "\t\t\t\t\t", $0;} /^Seats \t(Total: [0-9]+, Open: [0-9]+, Waitlist: [0-9]+)$/ {print "\t\t\t\t\t\t\t", $3, "\t", $6,  
"\t", $9;} /^[MWFTu]+ [0-9]+:[0-9]+/ {print "\t\t\t\t\t\t\t\t\t\t\t", $1, " ", $2, $3, $4;} /^[A-Z]+ [0-9][0-9][0-9][0-9]$/ {print  
"\t\t\t\t\t\t\t\t\t\t\t\t\t\t\t\t\t", $1, "\t", $3;}} | less
```

Or to try another way of formatting (still looks ugly though...)

```
cat cmsc.txt | awk -F'[ ,\|]' 'BEGIN {print "Course No.\tSection No.\tInstructor\t\tSeats\t\tOpen\t\tWaitlist\tDays  
Time\t\t\tBldg.\tRoom No."} /^CMSC[0-9][0-9][0-9]$/ {course = $1;} /^[0-9][0-9][0-9][0-9]$/ {section = $0;} /^[A-Za-z]+([A-Za-z]+)$*/ {instr = $0;} /^Seats \\\(Total: [0-9]+, Open: [0-9]+, Waitlist: [0-9]+\\)$/ {seats = $3; open = $6; waitlist = $9;}  
/^\\[MWFTuh]+[0-9]+:[0-9]+/ {days = $1; time1 = $2; time2 = $3; time3 = $4;} /^[A-Z]+ [0-9][0-9][0-9][0-9]$/ {bldg = $1;  
room = $3; print course, "\t", section, "\t\t", instr, "\t\t", seats, "\t\t\t", open, "\t\t\t\t", waitlist, "\t\t\t\t\t", days, "\t\t\t\t\t\t",  
time1, "\t\t\t\t\t\t", time2, "\t\t\t\t\t\t", time3;} | less
```

UNIX Tools Command: World Cup 1

```
cat worldcup.txt | sed 's/FIFA//g' | sed 's/[{}]\|\\(\|\\)\|/g' | sed 's/fb//g' | sed 's/style=\"background:#fff68f\"//g' | sed 's/<sup>//g' | sed 's/<\sup>//g' | sed 's/\[\#[0-9]\.\|\\)\|/g' | sed 's/style=white-space:nowrap//g' | awk -F[, \[\]] 'BEGIN {place = 0;} /[A-Z][A-Z][A-Z]/ { coun = $1; } /[0-9] \|[[([0-9][0-9][0-9][0-9])]*]/ { if (place == 4) place = 1; else place = place + 1; if (length($4) != 0) print coun, $4, place; if (length($11) != 0) print coun, $11, place; if (length($18) != 0) print coun, $18, place; if (length($25) != 0) print coun, $25, place; if (length($32) != 0) print coun, $32, place;}' | less
```

Python Script: CMSC

See `cmisc.py` for source code

| | | Course No. | Section No. | Instructor | Seats | Open | Waitlist | Days | Time | Building | Room |
|---|---|------------|-------------|----------------|-------|------|----------|------|------------------|----------|------|
| 0 | 0 | CMSC100 | 0101 | Charles Kassir | 45 | 4 | 0 | M | 4:00pm - 4:50pm | CSI | 2117 |
| 1 | 0 | CMSC106 | 0101 | Jianwu Wang | 45 | 0 | 5 | TuTh | 9:30am - 10:45am | CSI | 2117 |
| 2 | 0 | CMSC131 | 0101 | Evan Golub | 31 | 0 | 0 | MWF | 2:00pm - 2:50pm | CSI | 1115 |
| | 1 | CMSC131 | 0102 | Evan Golub | 31 | 0 | 0 | MWF | 2:00pm - 2:50pm | CSI | 1115 |
| | 2 | CMSC131 | 0103 | Evan Golub | 31 | 1 | 0 | MWF | 2:00pm - 2:50pm | CSI | 1115 |

Python Script: World Cup 1

| | | Country | Place | Year |
|---|---|---------|-------|------|
| 0 | 0 | BRA | 1 | 1958 |
| | 1 | BRA | 1 | 1962 |
| | 2 | BRA | 1 | 1970 |
| | 3 | BRA | 1 | 1994 |
| | 4 | BRA | 1 | 2002 |

See [worldcup.py](#) for source code

Python Script: World Cup 2

See worldcup.py for source code (last few lines in the file are for World Cup 2)

| Year | 1930 | 1934 | 1938 | 1950 | 1954 | 1958 | 1962 | 1966 | 1970 | 1974 | 1978 | 1982 | 1986 | 1990 | 1994 | 1998 | 2002 | 2006 | 2010 | 2014 |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Country | | | | | | | | | | | | | | | | | | | | |
| ARG | 2 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 1 | NaN | 1 | 2 | NaN | NaN | NaN | NaN | NaN | 2 |
| AUT | NaN | 2 | NaN | NaN | 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| BEL | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| BRA | NaN | NaN | 3 | 2 | NaN | 1 | 1 | NaN | 1 | 4 | 3 | NaN | NaN | NaN | 1 | 2 | 1 | NaN | NaN | 4 |
| BUL | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 1 | NaN | NaN | NaN | NaN | NaN |
| CHI | NaN | NaN | NaN | NaN | NaN | NaN | 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| CRO | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 1 | NaN | NaN | NaN | NaN |
| ENG | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 1 | NaN | NaN | NaN | NaN | NaN | 2 | NaN | NaN | NaN | NaN | NaN | NaN |