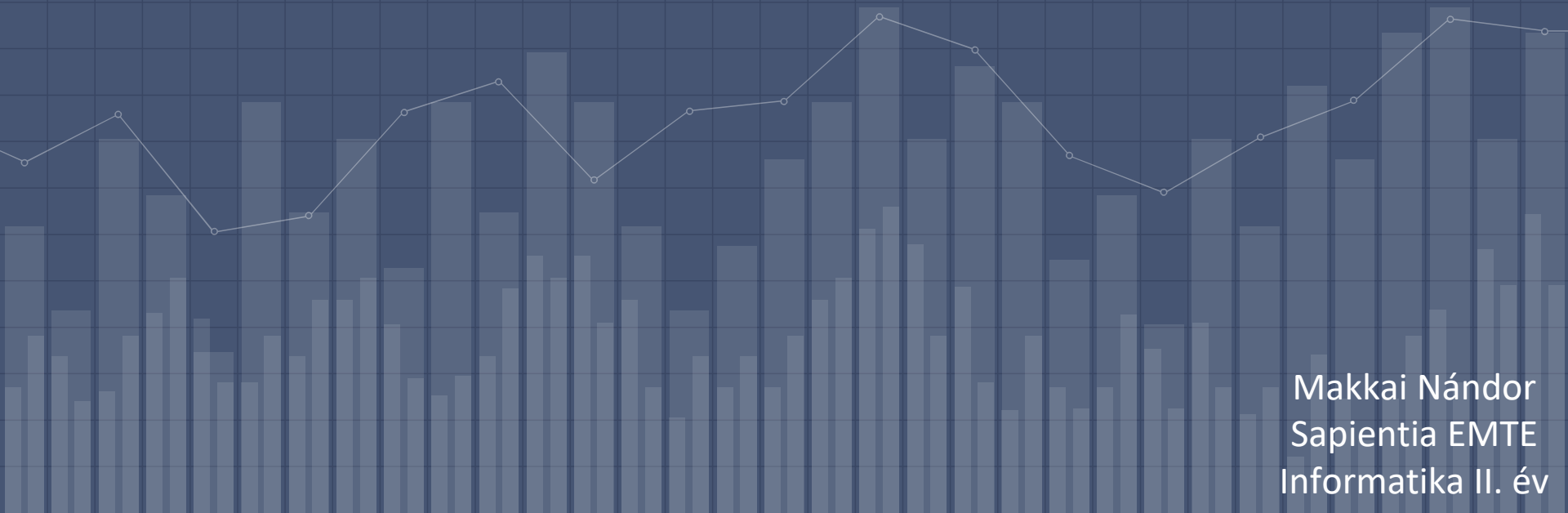


# Csíkszereda a középpontban



Makkai Nándor  
Sapientia EMTE  
Informatika II. év

# Az adathalmaz

## Klímaadatok: mérések, állomások

### Problémák:

- Összeragadt értékek
- Csillagok a hiányzó adatok helyén
- Adatok beimportálása adatbázisba
- Az esetleges duplikátumok kiszűrése
- Rossz mérési adatok

### Megoldások:

- A .txt fájlokat .csv fájlkká alakítani, úgy hogy az összeragadt értékeket szétválasztjuk és a csillagokat kiszedjük
- Adatokat beimportálni adatbázisba, duplikátumokat kiszűrni
- Rossz mérési adatokat kiküszöbölni

# Fájlok átalakítása

- Program C nyelven
- txt → csv
- Új fájlt hoz létre
- Soronként olvassa a txt-t
- Szétválasztja az összeragadt adatokat (0.00T)
- Kiszedi a csillagokat

```
1 #include <stdio.h>
2 #include <string.h>
3
4 int main(){
5     FILE* fin;
6     FILE* fout;
7     fin = fopen("ro\\9812808043220\\9812808043220dat.txt","r");
8     fout = fopen("ro\\9812808043220\\98.csv","w");
9     if(fin == NULL){
10         printf("No input file!");
11         return 1;
12     }
13     if(fout == NULL){
14         printf("No output file!");
15         return 1;
16     }
17     char line[150];
18     char delim[] = " ";
19     char search[] = "0.00T";
20     int size = strlen(search);
21     int pos;
22     char* ptr;
23     while(fscanf(fin,"%[^\\n\\n],line) != EOF){
24         ptr = strstr(line,search);
25         while(ptr != NULL){
26             pos = ptr - line;
27             for(int i = pos ; i < pos+size ; ++i){
28                 line[i-1] = line[i];
29             }
30             line[pos+size-1] = ' ';
31             ptr = strstr(line+pos+size,search);
32         }
33         ptr = strtok(line,delim);
34         if(ptr[0] != '*')
35             fprintf(fout,"%s",ptr);
36         while(1){
37             ptr = strtok(NULL,delim);
38             if(ptr == NULL)
39                 break;
40             if(ptr[0] == '*')
41                 fprintf(fout,",");
42             else
43                 fprintf(fout,"%s",ptr);
44         }
45         fprintf(fout,"\\n");
46     }
47     fclose(fin);
48     fclose(fout);
49     return 0;
50 }
```

# txt -> csv (stn fájl)

9812808043220stn - Jegyzetömb

Fájl Szerkesztés Formátum Nézet Súgó

USAF-WBAN_ID	STATION NAME	COUNTRY	STATE	LATITUDE	LONGITUDE	ELEVATION
150000	99999	DARABANI	ROMANIA	+48.200	+026.583	+0260.0
150002	99999	BORCEA FETESTI AIR BASE	ROMANIA	+44.383	+027.717	+0054.0
150040	99999	SIGHETUL MARMATIEI	ROMANIA	+47.933	+023.900	+0276.0
150070	99999	RADAUTI	ROMANIA	+47.833	+025.900	+0391.0
150090	99999	DOROHOI	ROMANIA	+47.967	+026.400	+0241.0
150100	99999	SATU MARE	ROMANIA	+47.703	+022.886	+0126.2
150140	99999	TAUTII MAGHERAUS	ROMANIA	+47.658	+023.470	+0184.1
150150	99999	OCNA SUGATAG	ROMANIA			
150200	99999	BOTOSANI	ROMANIA			
150230	99999	STEFAN CEL MARE	ROMANIA			
150235	99999	SUCEAVA/SALCEA	ROMANIA			
150250	99999	STANCA STEFANESTI	ROMANIA			
150320	99999	RODNEI MOUNTAIN RNG	ROMANIA			
150330	99999	IEZER	ROMANIA			
150400	99999	CIMPULUNG MOLDOVENE	ROMANIA			
150410	99999	RAUSENI	ROMANIA			
150420	99999	SACUIENI	ROMANIA			
150440	99999	SUPURU DE JOS	ROMANIA			
150470	99999	TARGU LAPUS	ROMANIA			
150520	99999	RARAU	ROMANIA			
150550	99999	RADASENI	ROMANIA			
150560	99999	COTNARI	ROMANIA			

stn - Jegyzetömb

Fájl Szerkesztés Formátum Nézet Súgó

```
USAF,WBAN_ID,STATION NAME,COUNTRY,LATITUDE, LONGITUDE,ELEVATION
150000,99999,DARABANI,ROMANIA,48.200,026.583,0260.0
150002,99999,BORCEA FETESTI AIR BASE,ROMANIA,44.383,027.717,0054.0
150040,99999,SIGHETUL MARMATIEI,ROMANIA,47.933,023.900,0276.0
150070,99999,RADAUTI,ROMANIA,47.833,025.900,0391.0
150090,99999,DOROHOI,ROMANIA,47.967,026.400,0241.0
150100,99999,SATU MARE,ROMANIA,47.703,022.886,0126.2
150140,99999,TAUTII MAGHERAUS,ROMANIA,47.658,023.470,0184.1
150150,99999,OCNA SUGATAG,ROMANIA,47.733,023.950,0503.0
150200,99999,BOTOSANI,ROMANIA,47.733,026.650,0163.0
150230,99999,STEFAN CEL MARE,ROMANIA,47.688,026.354,0419.1
150235,99999,SUCEAVA/SALCEA,ROMANIA,47.683,026.350,0399.0
150250,99999,STANCA STEFANESTI,ROMANIA,47.833,027.217,0110.0
150320,99999,RODNEI MOUNTAIN RNG,ROMANIA,47.583,024.667,2303.0
150330,99999,IEZER,ROMANIA,47.600,024.650,1786.0
150400,99999,CIMPULUNG MOLDOVENE,ROMANIA,47.533,025.567,0660.0
150410,99999,RAUSENI,ROMANIA,47.567,027.200,0061.0
150420,99999,SACUIENI,ROMANIA,47.350,022.100,0113.0
150440,99999,SUPURU DE JOS,ROMANIA,47.450,022.783,0161.0
150470,99999,TARGU LAPUS,ROMANIA,47.433,023.867,0365.0
150520,99999,RARAU,ROMANIA,47.450,025.567,1574.0
150550,99999,RADASENI,ROMANIA,47.467,026.250,0230.0
150560,99999,COTNARI,ROMANIA,47.367,026.933,0289.0
```

# txt -> csv (dat fájl)

9812808043220dat - Jegyzetömb

Fájl Szerkesztés Formátum Nézet Súgó

153770	99999	197511220600	070	7	***	14	OVC	6	*	0.5	46	*****	2	36	36	1013.1	*****	36	*****	0
153770	99999	197511220900	090	7	***	7	OVC	0	2	0.5	73	*****	7	32	32	1014.4	*****	*****	*****	*****
153770	99999	197511221200	070	11	***	***	OVC	0	0	1.2	71	*****	7	34	34	1013.3	*****	0.16	*****	*****
153770	99999	197511221800	050	9	***	14	OVC	7	2	3.8	71	*****	7	32	32	1013.7	*****	36	*****	*****
153770	99999	197511230000	050	4	***	14	OVC	7	2	2.5	71	*****	7	32	32	1013.7	*****	0.08	*****	*****
153770	99999	197511230300	050	11	***	14	OVC	7	2	5.0	71	*****	7	30	30	1013.9	*****	*****	*****	*****
153770	99999	197511230600	050	13	***	14	OVC	7	2	9.4	02	*****	7	30	25	1013.6	*****	30	*****	*****
153770	99999	197511231200	050	13	***	39	OVC	7	2	5.0	10	*****	7	32	28	1013.5	*****	0.00T	*****	*****
153770	99999	197511231800	020	2	***	14	OVC	7	2	9.4	71	*****	7	32	32	1015.1	*****	36	*****	*****
153770	99999	197511232100	360	16	***	***	OVC	*	*	7.5	71	*****	7	30	30	*****	*****	*****	*****	*****
153770	99999	197511240000	***	0	***	14	OVC	7	2	9.4	02	*****	7	28	27	*****	*****	*****	*****	*****
153770	99999	197511240300	***	0	***	14	OVC	7	2	9.4	02	*****	2	27	27	*****	*****	*****	*****	*****
153770	99999	197511240600	250	11	***	14	OVC	7	2	5.0	71	*****	2	28	27	*****	*****	*****	*****	*****
153770	99999	197511241800	020	7	***	14	OVC	7	2	5.0	71	*****	7	27	27	*****	*****	*****	*****	*****
153770	99999	197511250000	360	2	***	722	SCT	0	3	0	9.4	01	*****	7	19	19	*****	*****	*****	*****
153770	99999	197511250600	340	2	***	118	OVC	5	7	12.5	03	*****	1	19	19	*****	*****	*****	*****	*****
153770	99999	197511251500	320	7	***	14	OVC	7	2	1.2	10	*****	7	27	27	*****	*****	*****	*****	*****
153770	99999	197511251800	***	0	***	14	OVC	7	2	7.5	03	*****	7	27	27	*****	*****	*****	*****	*****
153770	99999	197511260000	***	0	***	25	OVC	5	*	7.5	02	*****	6	25	25	*****	*****	*****	*****	*****
153770	99999	197511260300	***	0	***	1	OVC	5	0	4	7.5	02	*****	2	23	23	*****	*****	*****	*****
153770	99999	197511260600	340	2	***	25	OVC	5	*	5.0	10	*****	2	27	27	*****	*****	*****	*****	*****
153770	99999	197511260900	070	4	***	14	OVC	0	2	3.8	10	*****	2	27	27	*****	*****	*****	*****	*****
153770	99999	197511261200	090	2	***	14	OVC	0	2	2.5	10	*****	7	28	28	*****	*****	*****	*****	*****
153770	99999	197511261500	***	0	***	14	OVC	6	*	0.5	46	*****	2	23	23	*****	*****	*****	*****	*****
153770	99999	197511261800	***	0	***	14	OVC	6	*	0.5	44	*****	4	25	25	*****	*****	*****	*****	*****

98 - Jegyzetömb

Fájl Szerkesztés Formátum Nézet Súgó

153770	99999	197511220600	070	7	***	14	OVC	6	*	0.5	46	*****	2	36	36	1013.1	*****	36	*****
153770	99999	197511220900	090	7	***	7	OVC	0	2	0.5	73	*****	7	32	32	1014.4	*****	*****	*****
153770	99999	197511221200	070	11	***	***	OVC	0	0	1.2	71	*****	7	34	34	1013.3	*****	0.16	*****
153770	99999	197511221800	050	9	***	14	OVC	7	2	3.8	71	*****	7	32	32	1013.7	*****	36	*****
153770	99999	197511230000	050	4	***	14	OVC	7	2	2.5	71	*****	7	32	32	1013.7	*****	0.08	*****
153770	99999	197511230300	050	11	***	14	OVC	7	2	5.0	71	*****	7	30	30	1013.9	*****	*****	*****
153770	99999	197511230600	050	13	***	14	OVC	7	2	9.4	02	*****	7	30	25	1013.6	*****	30	*****
153770	99999	197511231200	050	13	***	39	OVC	7	2	5.0	10	*****	7	32	28	1013.5	*****	0.00T	*****
153770	99999	197511231800	020	2	***	14	OVC	7	2	9.4	71	*****	7	32	32	1015.1	*****	36	*****
153770	99999	197511232100	360	16	***	***	OVC	*	*	7.5	71	*****	7	30	30	*****	*****	*****	*****
153770	99999	197511240000	***	0	***	14	OVC	7	2	9.4	02	*****	7	28	27	*****	*****	*****	*****
153770	99999	197511240300	***	0	***	14	OVC	7	2	9.4	02	*****	2	27	27	*****	*****	*****	*****
153770	99999	197511240600	250	11	***	14	OVC	7	2	5.0	71	*****	2	28	27	*****	*****	*****	*****
153770	99999	197511241800	020	7	***	14	OVC	7	2	5.0	71	*****	7	27	27	*****	*****	*****	*****
153770	99999	197511250000	360	2	***	722	SCT	0	3	0	9.4	01	*****	7	19	19	*****	*****	*****
153770	99999	197511250600	340	2	***	118	OVC	5	7	12.5	03	*****	1	19	19	*****	*****	*****	*****
153770	99999	197511251500	320	7	***	14	OVC	7	2	1.2	10	*****	7	27	27	*****	*****	*****	*****
153770	99999	197511251800	***	0	***	14	OVC	7	2	7.5	03	*****	7	27	27	*****	*****	*****	*****
153770	99999	197511260000	***	0	***	25	OVC	5	*	7.5	02	*****	6	25	25	*****	*****	*****	*****
153770	99999	197511260300	***	0	***	1	OVC	5	0	4	7.5	02	*****	2	23	23	*****	*****	*****
153770	99999	197511260600	340	2	***	25	OVC	5	*	5.0	10	*****	2	27	27	*****	*****	*****	*****
153770	99999	197511260900	070	4	***	14	OVC	0	2	3.8	10	*****	2	27	27	*****	*****	*****	*****
153770	99999	197511261200	090	2	***	14	OVC	0	2	2.5	10	*****	7	28	28	*****	*****	*****	*****
153770	99999	197511261500	***	0	***	14	OVC	6	*	0.5	46	*****	2	23	23	*****	*****	*****	*****
153770	99999	197511261800	***	0	***	14	OVC	6	*	0.5	44	*****	4	25	25	*****	*****	*****	*****

# Importálás

- Összesen 38.904.121 mérés
- Kezdetben ORACLE-be -> lassú
- Váltás SQLite-ra
- Két tábla: ClimateDB és Stations
- Elsődleges kulcsok: USAF és dátum, illetve USAF (duplikátumok kiszűrése)
- ClimateDB: 38.900.402 adat
- > 3.719 duplikátum
- Stations: 258 állomás

```

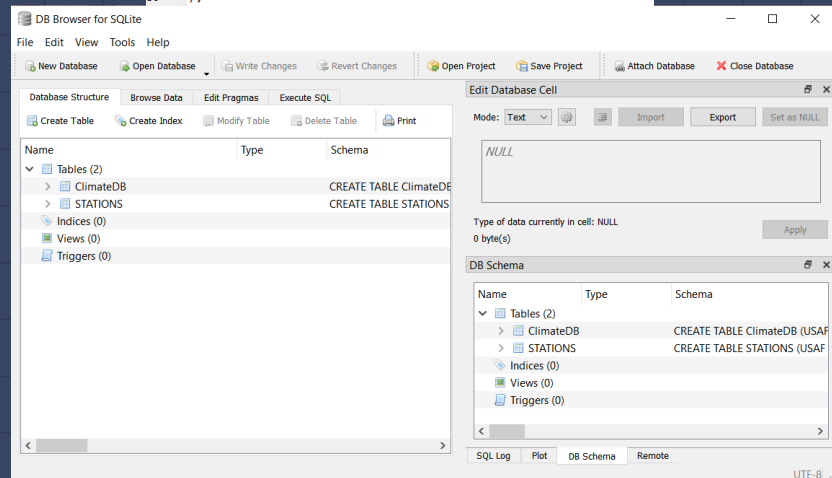
1 CREATE TABLE "ClimateDB" (
2   "USAF" VARCHAR NOT NULL,
3   "WBAN" VARCHAR NOT NULL,
4   "YRMODAHRMN" VARCHAR NOT NULL,
5   "DIR" VARCHAR,
6   "SPD" INTEGER,
7   "GUS" INTEGER,
8   "CLG" INTEGER,
9   "SRC" VARCHAR,
10  "I" INTEGER,
11  "M" INTEGER,
12  "H" INTEGER,
13  "VSB" REAL,
14  "MW1" VARCHAR,
15  "MW2" VARCHAR,
16  "MW3" VARCHAR,
17  "MW4" VARCHAR,
18  "AW1" VARCHAR,
19  "AW2" VARCHAR,
20  "AW3" VARCHAR,
21  "AW4" VARCHAR,
22  "W" INTEGER,
23  "TEMP" INTEGER,
24  "DEWP" INTEGER,
25  "SLP" REAL,
26  "ALT" REAL,
27  "STP" REAL,
28  "MAX" INTEGER,
29  "MIN" INTEGER,
30  "PCP01" VARCHAR,
31  "PCP06" VARCHAR,
32  "PCP24" VARCHAR,
33  "PCPXX" VARCHAR,
34  "SD" INTEGER,
35  CONSTRAINT "CLIMATE_PK" PRIMARY KEY ("USAF", "YRMODAHRMN")
36 );

```

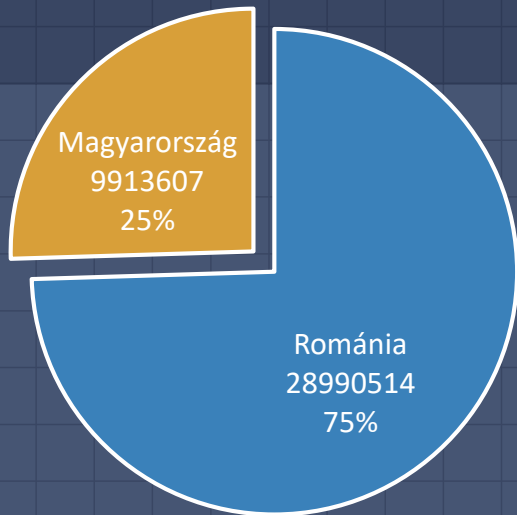
```

1 CREATE TABLE "STATIONS" (
2   "USAF" VARCHAR NOT NULL,
3   "WBAN" VARCHAR,
4   "STATION_NAME" VARCHAR,
5   "COUNTRY" VARCHAR,
6   "LATITUDE" VARCHAR,
7   "LONGITUDE" VARCHAR,
8   "ELEVATION" VARCHAR,
9   PRIMARY KEY ("USAF")
10 );

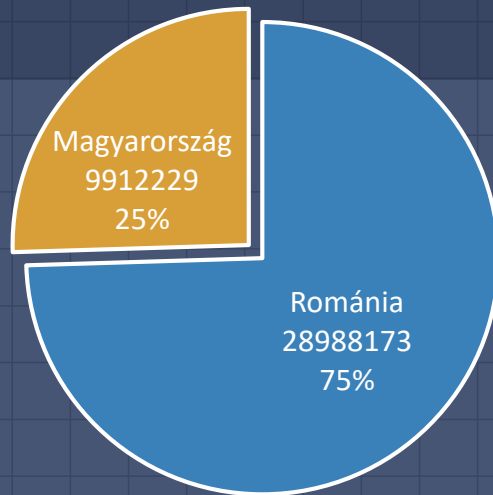
```



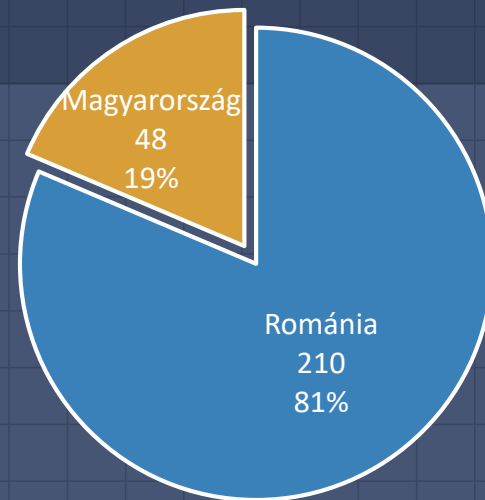
## Mérések



## Duplikátumok nélkül



## Állomások



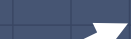
Ö: 38.900.402 adat  
258 állomás



150.776 adat / állomás



71 állomás,  
amelyek az  
átlagnál több  
méréssel  
rendelkeznek



30 magyarországi

41 romániai

M: 9.912.229 adat  
48 állomás



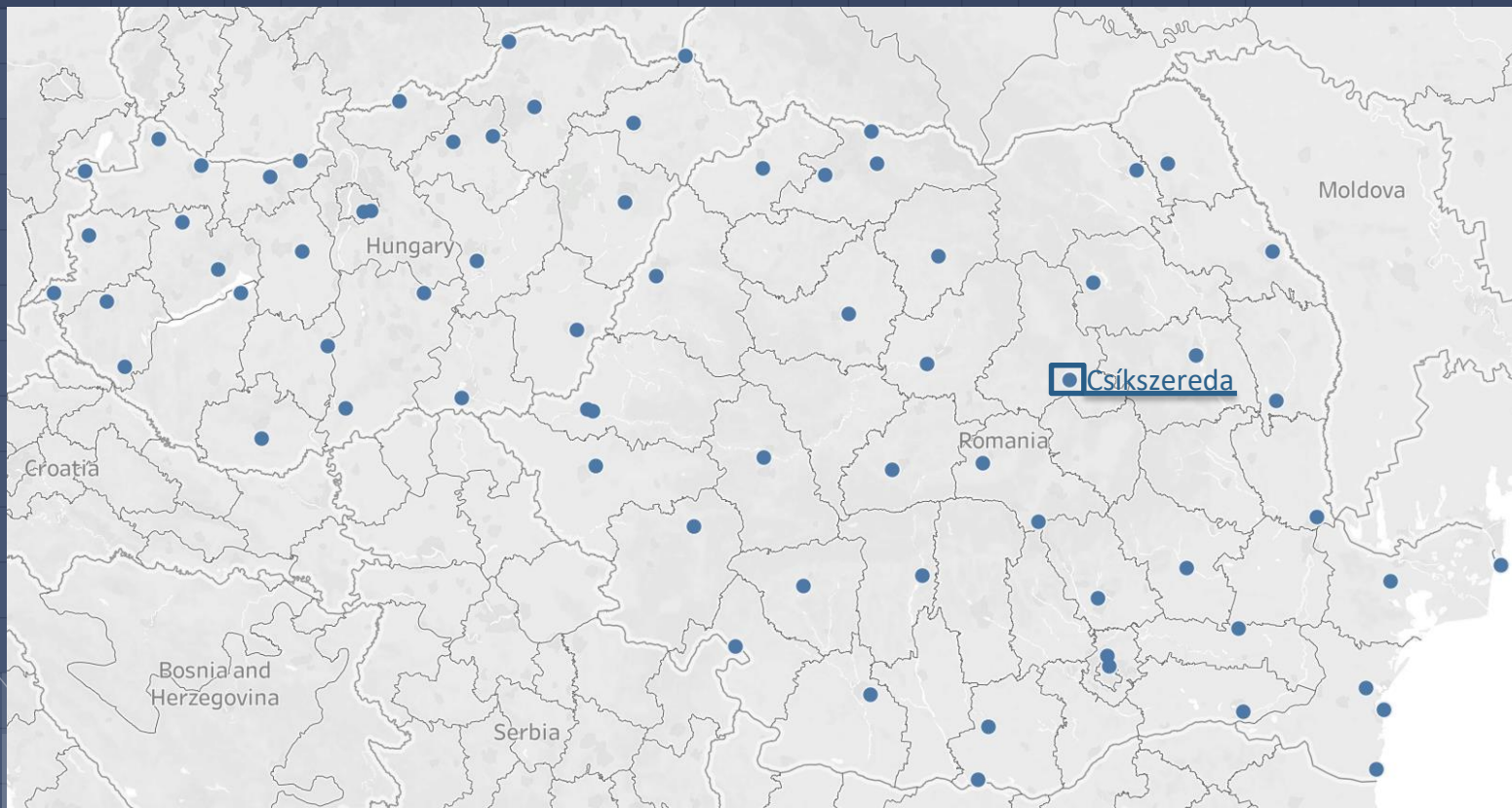
206.504 adat / állomás

R: 28.988.173 adat  
210 állomás



138.038 adat / állomás

# A 71 állomás (22.452.721 mérés)





# Rossz mérési adatok

- Extrém hőmérsékletek -38 °F és 113 °F
- Ami nem ebben az intervallumban van rossz mérés
- Ha DIR oszlop értéke NULL, akkor SPD oszlop értéke NULL vagy 0
- Ellenkező esetben a SPD értéke rossz mérés
- A VSB értéke nem lehet nagyobb 10-nél, ami ennél nagyobb visszaállítjuk 10-re

```
-- rossz homerseklet meresek kijavitasa
update ClimateDB
set TEMP = NULL
where temp < -38 or temp > 113;

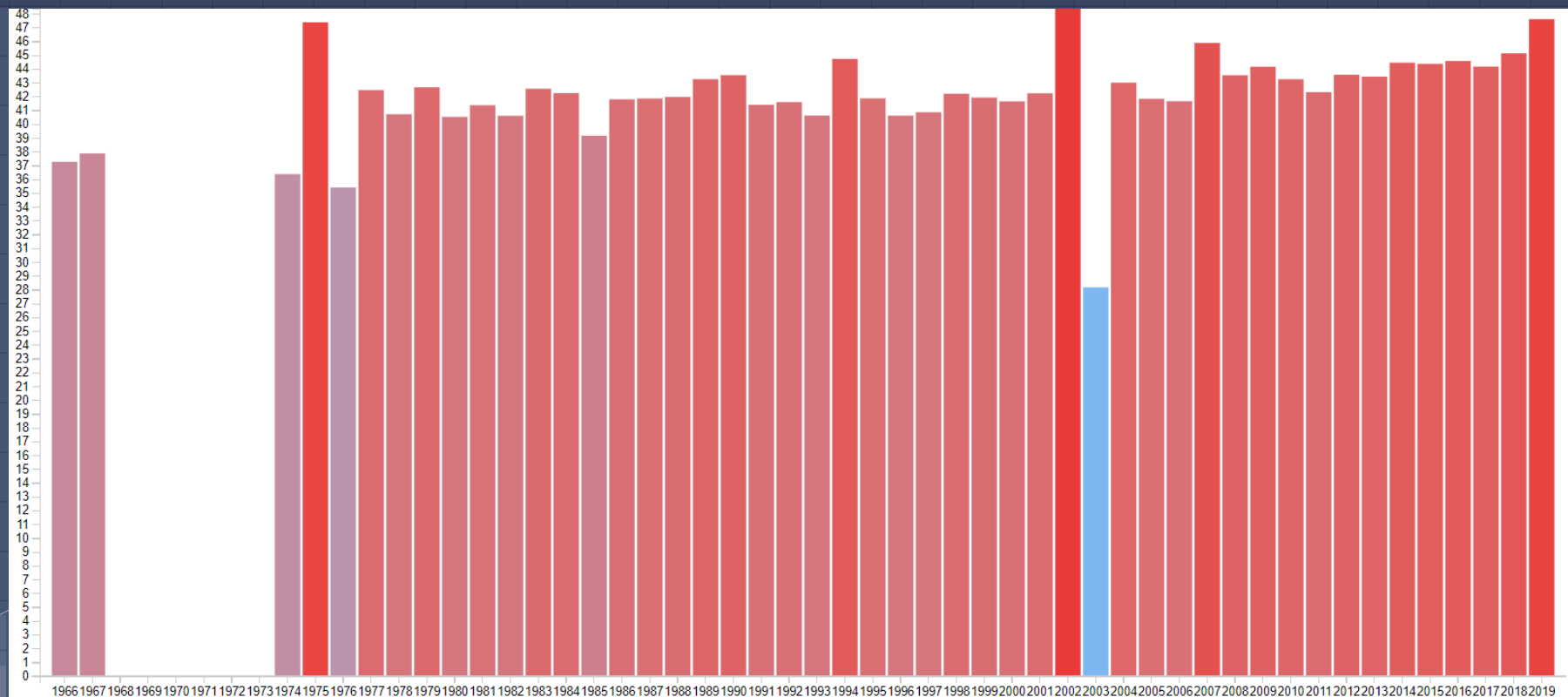
-- rossz szelsebesseg meresek kijavitasa
update ClimateDB
set SPD = NULL
where DIR is NULL
and SPD > 0;

-- rossz latotavolsagi meresek kijavitasa
update ClimateDB
set VSB = 10
where VSB > 10;
```

# Csíkszereda

- Hőmérséklet
- Szélsébség
- Köd
- Látótávolság
- Felhőzet

# Csíkszereda, átlaghőmérséklet évente (°F)



# Csíkszereda, átlaghőmérséklet évente (°F)

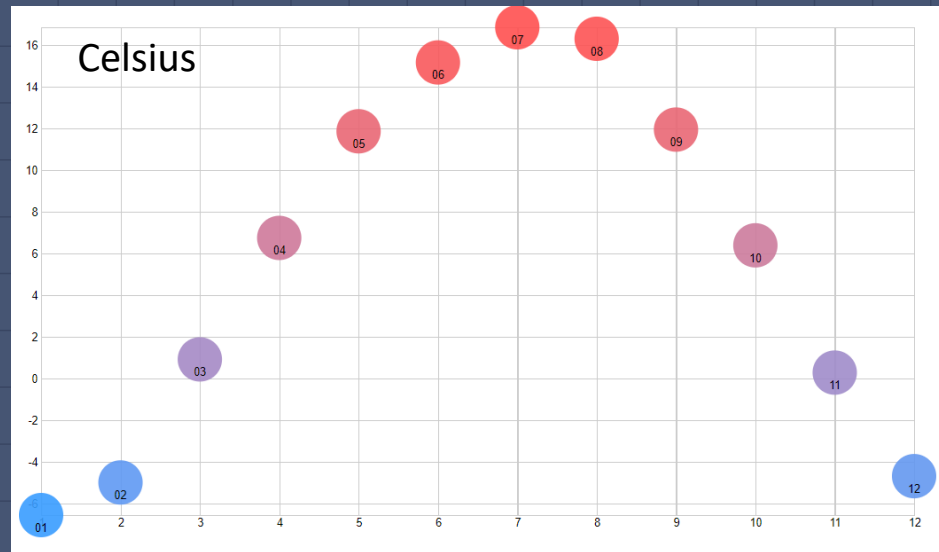
## SQL lekérdezés

- Csak azok a napok és hónapok, amikor volt legalább 5 hőmérséklet mérés

```
14 select substr(YRMODAHRMN,1,4) as year,  
15 avg(TEMP) as avg_F, (avg(TEMP)-32)/1.8 as avg_c  
16 from  
17 (select YRMODAHRMN, avg(TEMP) as temp  
18 from  
19 (select YRMODAHRMN, avg(TEMP) as temp  
20 from ClimateDB  
21 where USAF =  
22 (select USAF  
23 from stations  
24 where STATION_NAME = 'MIERCUREA CIUC'))  
25 group by substr(YRMODAHRMN,1,8)  
26 having count(TEMP) > 4)  
27 group by substr(YRMODAHRMN,1,6)  
28 having count(TEMP) > 4)  
29 group by substr(YRMODAHRMN,1,4)  
30 order by substr(YRMODAHRMN,1,4);
```

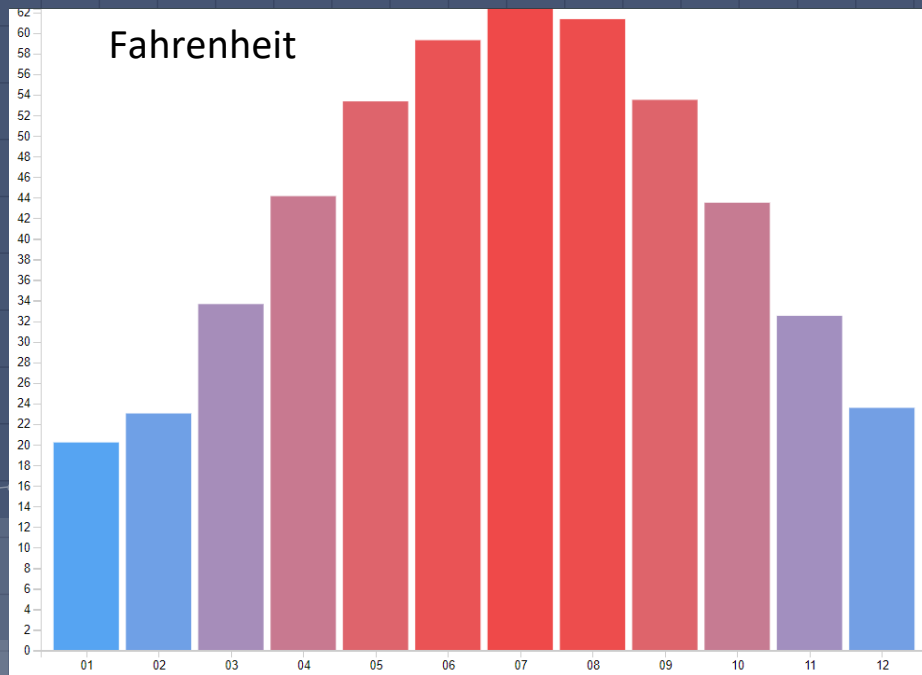
# Csíkszereda, átlaghőmérséklet hónaponta

Celsius



Január a leghidegebb és  
július a legmelegebb hónap

Fahrenheit



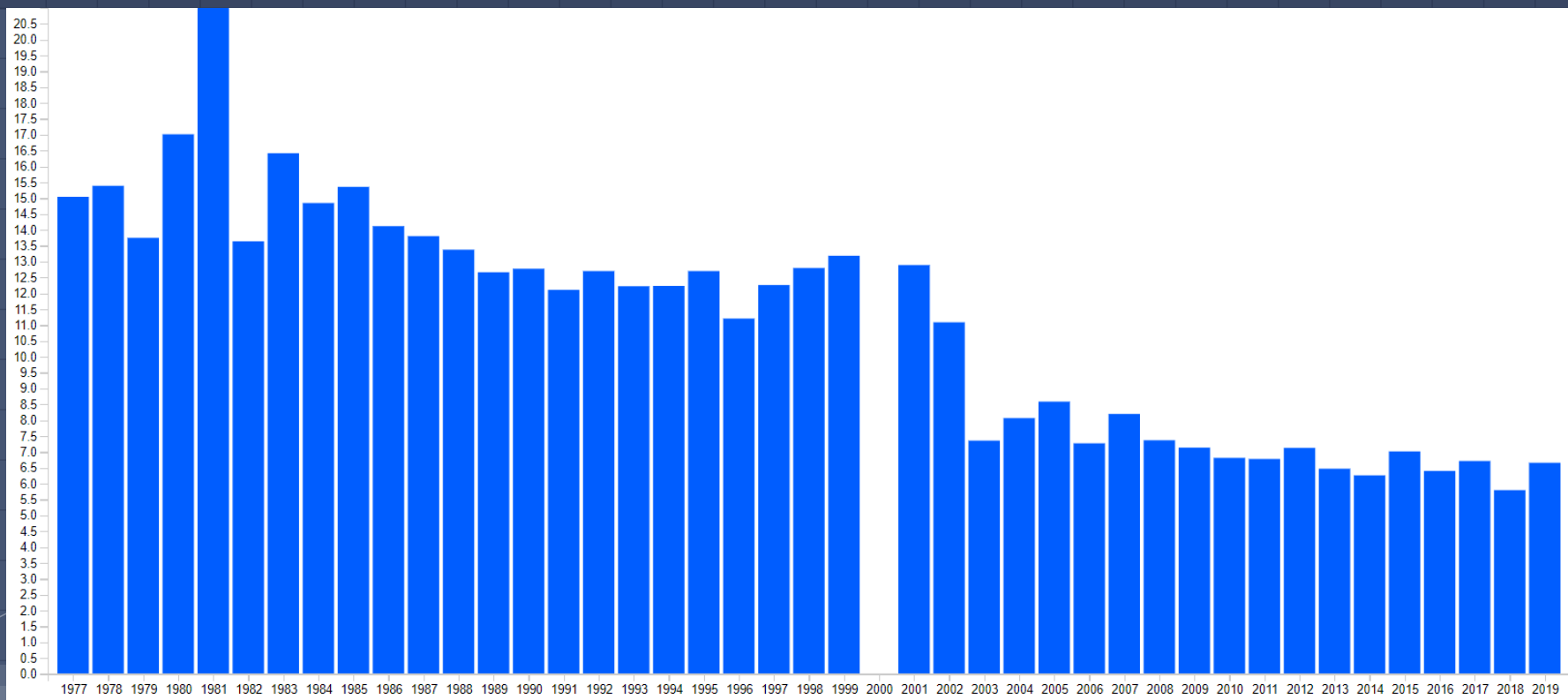
# Csíkszereda, átlaghőmérséklet hónaponta

## SQL lekérdezés

- Csak azok a napok és hónapok, amikor volt legalább 5 hőmérséklet mérés

```
34 select substr(YRMODAHRMN,5,2) as month,  
35 avg(TEMP) as avg_F, (avg(TEMP)-32)/1.8 as avg_c  
36 from  
37   (select YRMODAHRMN, avg(TEMP) as temp  
38     from  
39       (select YRMODAHRMN, avg(TEMP) as temp  
40         from ClimateDB  
41         where USAF =  
42           (select USAF  
43             from stations  
44             where STATION_NAME = 'MIERCUREA CIUC')  
45         group by substr(YRMODAHRMN,1,8)  
46         having count(TEMP) > 4)  
47       group by substr(YRMODAHRMN,1,6)  
48       having count(TEMP) > 4)  
49   group by substr(YRMODAHRMN,5,2)  
50   order by substr(YRMODAHRMN,5,2);
```

# Csíkszereda, átlagszélsősebesség évente (KPH)



# Csíkszereda, átlagszélsébség évente (KPH)

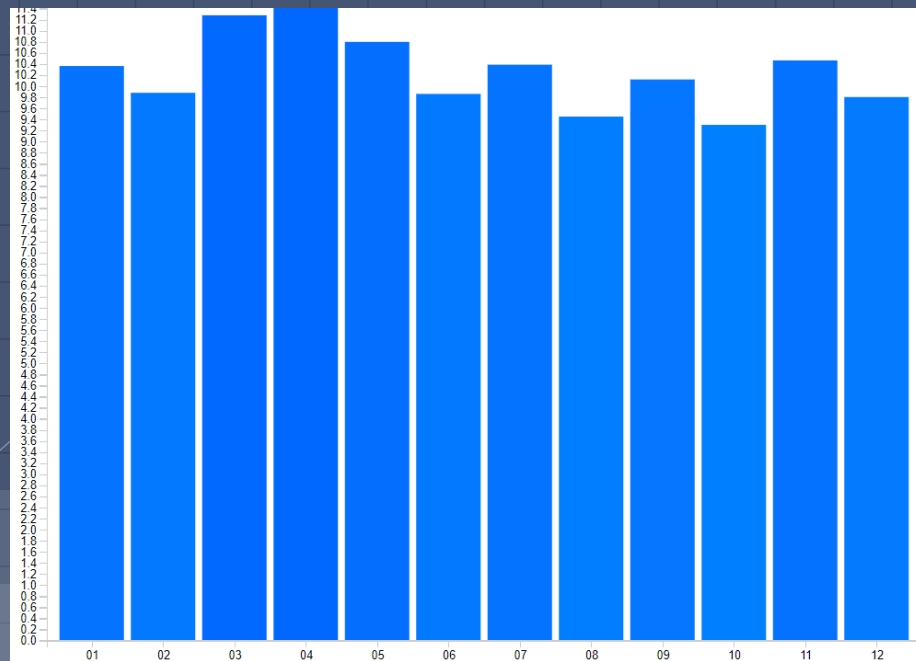
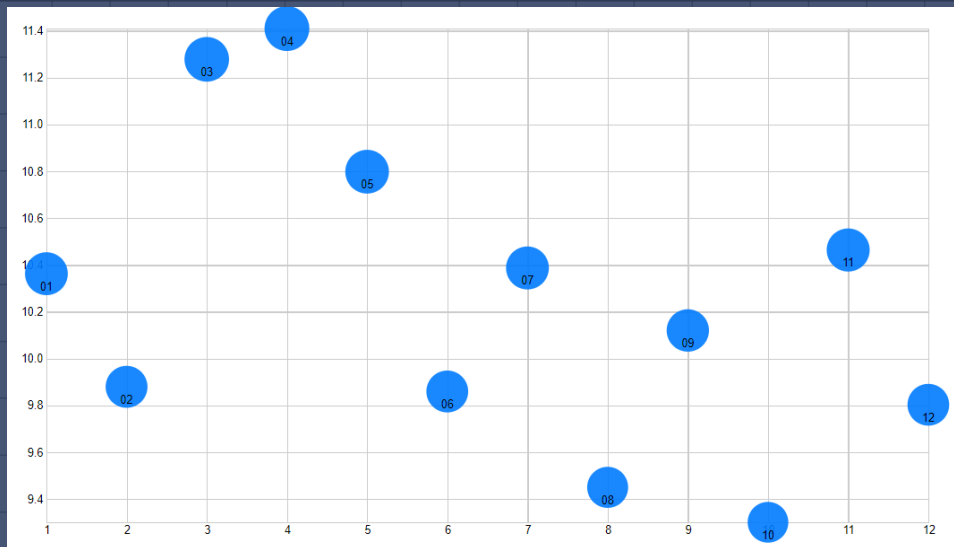
## SQL lekérdezés

- Csak azok a napok és hónapok, amikor volt legalább 5 szélsébség mérés

```
74 select substr(YRMODAHRMN,1,4) as year,  
75 avg(SPD) as avg_M, avg(SPD)*1.609344 as avg_K  
76 from  
77 (select YRMODAHRMN, avg(SPD) as spd  
78 from  
79 (select YRMODAHRMN, avg(SPD) as spd  
80 from ClimateDB  
81 where USAF =  
82 (select USAF  
83 from stations  
84 where STATION_NAME = 'MIERCUREA CIUC')  
85 and DIR is not NULL  
86 group by substr(YRMODAHRMN,1,8)  
87 having count(SPD) > 4)  
88 group by substr(YRMODAHRMN,1,6)  
89 having count(SPD) > 4)  
90 group by substr(YRMODAHRMN,1,4)  
91 order by substr(YRMODAHRMN,1,4);
```



# Csíkszereda, átlagszélebbesség hónaponta (KPH)



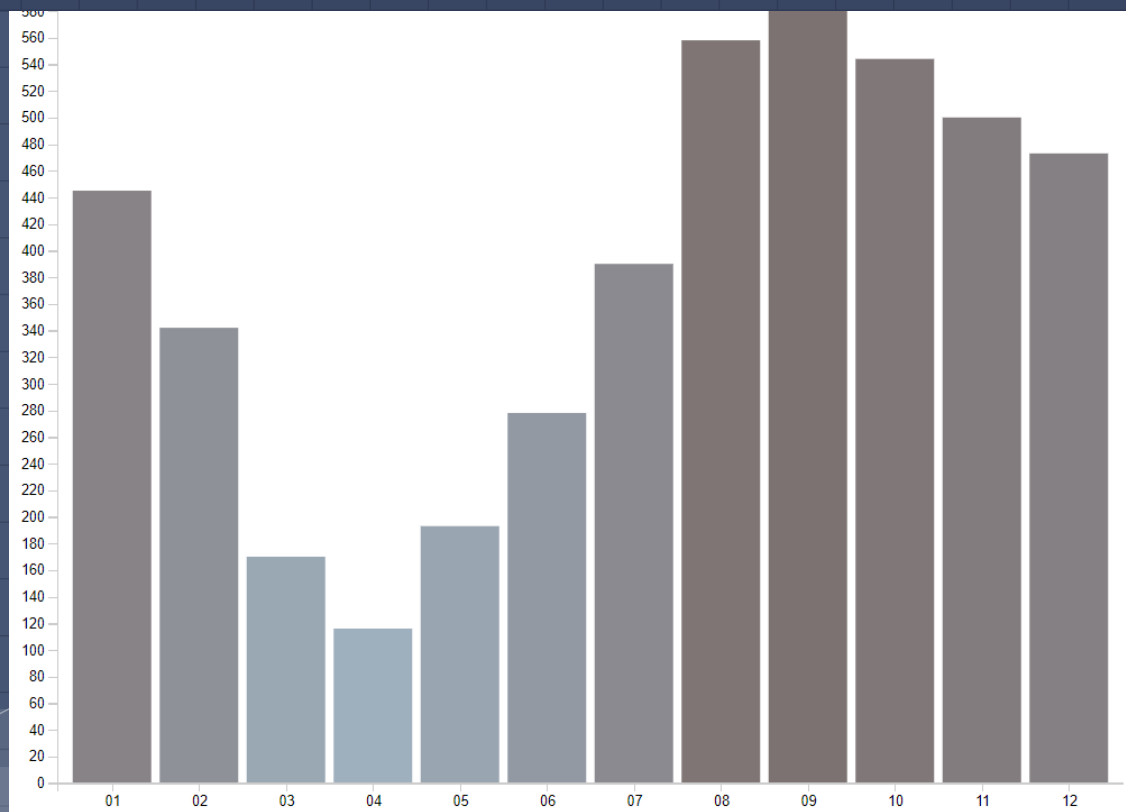
# Csíkszereda, átlagszélsébség hónaponta (KPH)

## SQL lekérdezés

- Csak azok a napok és hónapok, amikor volt legalább 5 szélsébség mérés

```
95 select substr(YRMODAHRMN,5,2) as month,  
96 avg(SPD) as avg_M, avg(spd)*1.609344 as avg_K  
97 from  
98 (select YRMODAHRMN, avg(SPD) as spd  
99 from  
100 (select YRMODAHRMN, avg(SPD) as spd  
101 from ClimateDB  
102 where USAF =  
103 (select USAF  
104 from stations  
105 where STATION_NAME = 'MIERCUREA CIUC')  
106 and DIR is not NULL  
107 group by substr(YRMODAHRMN,1,8)  
108 having count(SPD) > 4)  
109 group by substr(YRMODAHRMN,1,6)  
110 having count(SPD) > 4)  
111 group by substr(YRMODAHRMN,5,2)  
112 order by substr(YRMODAHRMN,5,2);
```

# Csíkszereda, köd mérések hónaponta



A legködösebb hónap  
szeptember, míg  
áprilisban jegyezték fel a  
legkevesebbszer ködöt

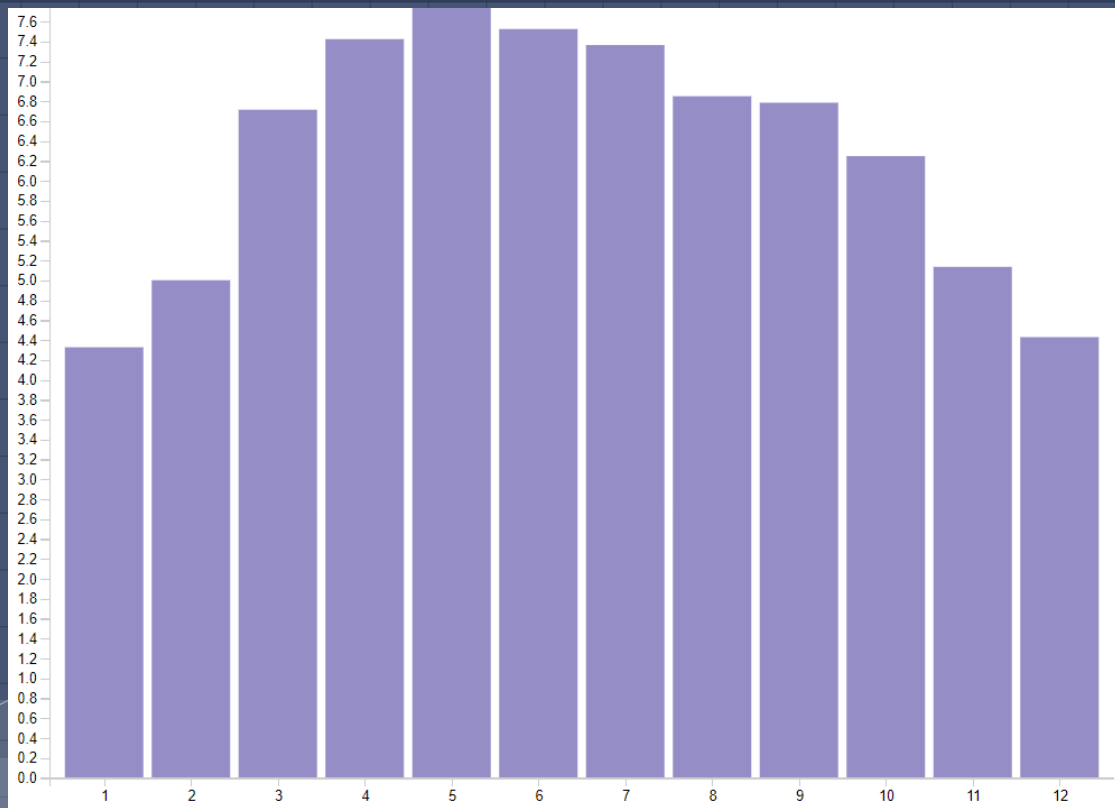
# Csíkszereda, kód mérések hónaponta

## SQL lekérdezés

- Megszámoljuk minden év minden hónapjában hány nap mérték ködöt, majd csoportosítjuk hónapokra és összeadjuk őket

```
55 select substr(YRMODAHRMN,5,2) as month, sum(kod) as kod
56 from
57 (select YRMODAHRMN, count(*) as kod
58  from
59  (select YRMODAHRMN
60   from ClimateDB
61   where USAF =
62    (select USAF
63     from STATIONS
64     where STATION_NAME = 'MIERCUREA CIUC')
65    and (MW1 between 40 and 49
66         or MW1 = 11
67         or MW1 = 12
68         or MW1 = 28
69         or AW1 between 40 and 49
70         or AW1 = 11
71         or AW1 = 12
72         or AW1 = 28
73         or W = 4)
74   group by substr(YRMODAHRMN,1,8))
75  group by substr(YRMODAHRMN,1,6))
76 group by substr(YRMODAHRMN,5,2)
77 order by substr(YRMODAHRMN,5,2);
```

# Csíkszereda, átlag látótávolság hónaponta (mérőföld)



Amelyik hónapban kevesebb alkalommal jegyezték fel ködöt, abban a hónapban nagyobb az átlag látótávolság

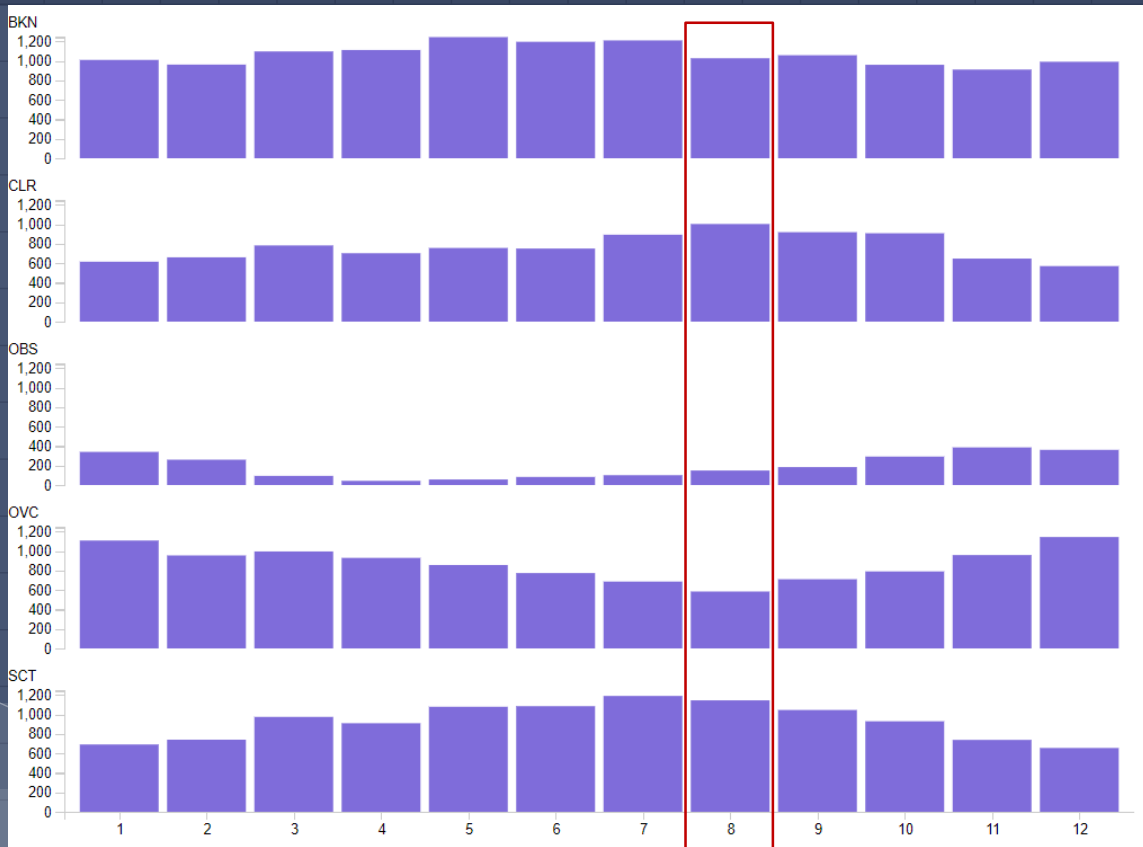
# Csíkszereda, átlag látótávolság hónaponta (mérőföld)

## SQL lekérdezés

- Minden napra kiszámoljuk az átlag látótávolságot, majd csoportosítjuk őket hónaponta és újra átlagot számolunk

```
45 select substr(datum,5,2) as honap, avg(avg_vsb) as avg_vsb
46 from
47 (select substr(YRMODAHRMN,1,8) as datum, avg(VSB) as avg_vsb
48 from ClimateDB
49 where USAF =
50 (select USAF
51 from STATIONS
52 where STATION_NAME = 'MIERCUREA CIUC')
53 group by substr(YRMODAHRMN,1,8))
54 group by substr(datum,5,2);
```

# Csíkszereda, felhőzet hónaponta



CLR – tiszta  
SCT – szétszórt, gyér  
BKN – egyenetlen  
OVC – felhős, borult  
OBS – borús

Augusztusban a legtisztább  
az ég Csíkszeredában

# Csíkszereda, felhőzet hónaponta

## SQL lekérdezés

- Megnézzük minden napra, hogy milyen felhőzeti feljegyzések vannak, majd csoportosítjuk hónapokra őket és megszámoljuk mindegyikből hány van

```
31 select substr(datum,5,2) as honap, SKC, count(*) as darab
32 from
33     (select substr(YRMODAHRMN,1,8) as datum, SKC
34      from ClimateDB
35      where USAF =
36           (select USAF
37            from STATIONS
38            where STATION_NAME = 'MIERCUREA CIUC')
39      and SKC is not NULL
40      group by substr(YRMODAHRMN,1,8), SKC)
41 group by substr(datum,5,2), SKC;
```



# Köszönöm a figyelmet!

Az adatvizualizációk a RAWGraphs segítségével készültek