

1.

```
> fastrunners=filter(TenMileRace,net<4000)
> view(fastrunners)
> nrow(fastrunners)
[1] 346
```

Environment History Connections Tutorial

Import Dataset 352 MIB List

R Global Environment

Data

fastrunners 346 obs. of 5 variables

\$ state: Factor w/ 62 levels "", "AK", "AL", "AR", ...: 20 43 31 25 43 61 11 57 24 31 ...

\$ time : int 3474 4165 3403 3540 3917 3917 3970 4300 2825 3410 ...

\$ net : int 3467 3897 3399 3533 3908 3875 3950 3971 2824 3404 ...

\$ age : int 17 18 19 19 19 19 19 20 20 ...

\$ sex : Factor w/ 2 levels "F", "M": 2 2 2 2 2 2 2 2 2 ...

Values

racetimes int [1:8636] 5978 4457 4928 4229 5076 5968 5928 5939 4077 4634 ...

Untitled1 x TenMileRace x fastrunners x

Filter

| | state | time | net | age | sex |
|----|-------|------|------|-----|-----|
| 1 | IL | 3474 | 3467 | 17 | M |
| 2 | NY | 4165 | 3897 | 18 | M |
| 3 | MD | 3403 | 3399 | 19 | M |
| 4 | KS | 3540 | 3533 | 19 | M |
| 5 | NY | 3917 | 3908 | 19 | M |
| 6 | WI | 3917 | 3875 | 19 | M |
| 7 | DC | 3970 | 3950 | 19 | M |
| 8 | VA | 4300 | 3971 | 19 | M |
| 9 | Kenya | 2825 | 2824 | 20 | M |
| 10 | MD | 3410 | 3404 | 20 | M |
| 11 | MD | 3413 | 3407 | 20 | M |

Showing 1 to 11 of 346 entries 5 total columns

2.

```
> femalerunners=filter(TenMileRace,sex=="F")
> view(femalerunners)
> nrow(femalerunners)
a. [1] 4325

> foreignrunners=filter(TenMileRace,stri_length(state)>2)
> view(foreignrunners)
> nrow(foreignrunners)
[1] 32
b. > |
```

3.

```
> table(TenMileRace$state)
```

| | | | | |
|------|------|---------|-------------|-----------|
| | AK | AL | AR | Australia |
| 1 | 1 | 2 | 1 | 1 |
| AZ | CA | CO | Colombia | CT |
| 3 | 23 | 7 | 1 | 35 |
| DC | DE | EN | Ethiopia | FL |
| 1642 | 20 | 1 | 2 | 20 |
| GA | GE | HI | IA | IL |
| 23 | 2 | 2 | 11 | 32 |
| IM | IN | Japan | Kenya | KS |
| 1 | 7 | 2 | 14 | 8 |
| KY | LA | LE | Lithuania | MA |
| 3 | 7 | 1 | 1 | 63 |
| MD | ME | MI | MN | MO |
| 2166 | 15 | 17 | 11 | 10 |
| NC | ND | NE | New Zealand | NH |
| 54 | 1 | 1 | 1 | 11 |
| NJ | NM | NY | OH | OR |
| 92 | 4 | 211 | 34 | 6 |
| PA | RI | Romania | Russia | SA |
| 273 | 1 | 3 | 3 | 2 |
| SC | TN | TX | Uhwiesen | Ukraine |
| 2 | 9 | 24 | 2 | 1 |
| Usa | VA | VI | VT | WA |
| 1 | 3689 | 2 | 9 | 4 |
| WI | WV | | | |
| 17 | 23 | | | |

4.

```
> femalefastrunners=filter(fastrunners,sex=="F")
```

```
> malerunners=filter(TenMileRace,sex=="M")
> view(malerunners)
```

```
> malerunnersFasterThanFastestFemalerunners=filter(malerunners,net<min(femalefastrunners$net))
> View(malerunnersFasterThanFastestFemalerunners)
> range(malerunnersFasterThanFastestFemalerunners$net)
[1] 2814 3120
```

```
> nrow(malerunnersFasterThanFastestFemalerunners)
[1] 21
```

5.

I have two solutions for question 5 that provide the same output histogram.

Solution 1:

```
> totalRunners=nrow(TenMileRace)
> toprunners=totalRunners*(1-0.05)
> bottomrunners=totalRunners*0.05

> typical_runners=filter(TenMileRace,rank(TenMileRace$net)<toprunners&rank(TenMileRace$net)>bottomrunners)

> hist(typical_runners$net)
```

Solution 2:

```
> sortedRacetimes=sort(racetimes)

> max=quantile(sortedRacetimes, 0.05)
> min=quantile(sortedRacetimes, 1-0.05)

> typicalRunners=sortedRacetimes[which(sortedRacetimes>=max & sortedRacetimes<=min)]
> hist(typicalRunners)
```

