

# Tidy Tuesday: Scottish Munros

August 20, 2025

Today's goal is to answer the three questions on the Tidy Tuesday's readme file (<https://github.com/rfordatascience/tidytuesday/tree/main/data/2025/2025-08-19>).

- How many peaks currently listed as Munros have always been included on the list?
- Which year saw the largest number of changes to the classification?
- Which Munro is the most remote?

## Prepare

### *Knowing the data*

From the readme:

A Munro is a Scottish mountain with an elevation of over 3,000 feet, whereas a Munro Top is a subsidiary summit of a Munro that also exceeds 3,000 feet in height but is not considered a distinct mountain in its own right. The most famous Munro is Ben Nevis.

The data and analysis will be about Scottish mountains and their classification statuses over time.

### *Data Preparation*

Hide

```
# Import libraries
library(tidyverse)
library(sqldf)
library(ggplot2)
```

Hide

```
# Import the data
munros <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/main/data/2025/2025-08-19/scottish_munros.csv')
munros
```

DoBIH_num...	Name	Height...
<chr>	<chr>	<dbl>
1	Ben Chonzie	931.00

DoBIH_num...	Name	Height...
<chr>	<chr>	<dbl>
17	Ben Vorlich	985.30
18	Stuc a' Chroin	973.00
32	Ben Lomond	973.70
26	Ben More	1174.00
27	Stob Binnein	1165.00
28	Stob Binnein - Stob Coire an Lochain	1068.00
39	Stob Binnein - Creag a' Bhragit	923.00
33	Stob Binnein - Meall na Dige	966.00
30	Cruach Ardrain	1045.90
1 10 of 604 rows   1 5 of 19 columns		
Previous 1 2 3 4 5 6 61 Next		

First off, the columns `DoBIH_number` and `Comments` can be removed, as neither are relevant to the goals presented. Additionally, the final row of the dataset can be dropped, as it appears that that row does not contain any data and is instead a text block that describes the source of the data.

Hide

```
# Drop the 2 columns
drops <- c('DoBIH_number', 'Comments')
munros <- munros[ , !(names(munros) %in% drops)]

# Remove the last row
munros <- munros[-nrow(munros),]
```

I will also rename the columns describing the years to the format “y1891”. This is because R does not work well with column names being entirely made up of integers.

Hide

```
# Rename every column to fit formatting
munros <- munros %>%
  rename(
    y1891 = "1891",
    y1921 = "1921",
    y1933 = "1933",
    y1953 = "1953",
    y1969 = "1969",
    y1974 = "1974",
    y1981 = "1981",
    y1984 = "1984",
    y1990 = "1990",
    y1997 = "1997",
    y2021 = "2021"
  )
```

Also replace NA values with a string value of “NA”.

Hide

```
munros[is.na(munros)] <- "NA"
```

Now make sure data looks all ok.

Hide

```
# View cleaned data
munros
```

Name	Height...	Height_ft	xcc
<chr>	<dbl>	<dbl>	<c
Ben Chonzie	931.00	3054	277
Ben Vorlich	985.30	3233	262
Stuc a' Chroin	973.00	3192	261
Ben Lomond	973.70	3195	236
Ben More	1174.00	3852	243
Stob Binnein	1165.00	3822	243
Stob Binnein - Stob Coire an Lochain	1068.00	3504	243
Stob Binnein - Creag a' Bhragit	923.00	3028	244
Stob Binnein - Meall na Dige	966.00	3169	245
Cruach Ardrain	1045.90	3431	240

1-10 of 603 rows | 1-5 of 16 columns

Previous123456...61Next

Lastly, check the data types.

```
sapply(munros, class)
```

	Name	Height_m	Height_ft	xcoord	ycoord	y1891	y1921	y1933
y1953	y1969	y1974	y1981	y1984				
"character"	"numeric"	"numeric"	"numeric"	"numeric"	"character"	"character"	"character"	
"character"	"character"	"character"	"character"	"character"	"character"			
	y1990	y1997	y2021					
"character"	"character"	"character"						

Now, onto answering the questions.

# Analysis

1) How many peaks currently listed as Munros have always been included on the list?

First, let's see how many mountains were listed as a Munro in the initial survey:

```
# Find the count of Munros in 1891
sqldf("SELECT COUNT(Name) AS 'Number of Munros in 1891' FROM munros
      WHERE y1891 = 'Munro'")
```

Number of Munros in 1891	
	<int>
	283
1 row	

So 283 formations were classed as a Munro in the initial 1891 survey. Now let us see how many kept their Munro statues throughout the years:

Hide

```
# Find the count of Munros that kept a classification of "Munro" in every survey
sqldf("SELECT COUNT(Name) AS 'Forever Munros' FROM munros
      WHERE
        y1891 = 'Munro' AND
        y1921 = 'Munro' AND
        y1933 = 'Munro' AND
        y1953 = 'Munro' AND
        y1969 = 'Munro' AND
        y1974 = 'Munro' AND
        y1981 = 'Munro' AND
        y1984 = 'Munro' AND
        y1990 = 'Munro' AND
        y1997 = 'Munro' AND
        y2021 = 'Munro'")
```

Forever Munros
<int>
243
1 row

So out of the 283 Munros that were classed as such, **243** have kept their classification as a Munro in every survey between the first of 1891 to the most recent one of 2021.

For Munro Tops...

Hide

```
# Find the count of Munro Tops in 1891
sqldf("SELECT COUNT(Name) AS 'Number of Munro Tops in 1891' FROM munros
      WHERE y1891 = 'Munro Top'")
```

Number of Munro Tops in 1891
<int>
255
1 row

Hide

```
# Find the count of Munros that kept a classification of "Munro" in every survey
sqldf("SELECT COUNT(Name) AS 'Forever Munro Tops' FROM munros
WHERE
  y1891 = 'Munro Top' AND
  y1921 = 'Munro Top' AND
  y1933 = 'Munro Top' AND
  y1953 = 'Munro Top' AND
  y1969 = 'Munro Top' AND
  y1974 = 'Munro Top' AND
  y1981 = 'Munro Top' AND
  y1984 = 'Munro Top' AND
  y1990 = 'Munro Top' AND
  y1997 = 'Munro Top' AND
  y2021 = 'Munro Top'")
```

**Forever Munro Tops**

&lt;int&gt;

156

1 row

Additionally, 156 Munro Tops have consistently been classified as a such during the same surveys. Compared to the count of 255 in the initial 1891 survey.

## 2) Which year saw the largest number of changes to the classification?

There are 10 surveys following the first one in 1891. Thus, there is a possible 10 instances of Munro classification being updated between the years of Munro, Munro Top, and NA (not a Munro).

Hide

```
cat('1921 updates:', sum(!(munros$y1891 == munros$y1921)) )
```

1921 updates: 72

Hide

```
cat('\n')
```

Hide

```
cat('1933 updates:', sum(!(munros$y1921 == munros$y1933)) )
```

1933 updates: 6

Hide

```
cat('\n')
```

Hide

```
cat('1953 updates:', sum(!(munros$y1933 == munros$y1953)) )
```

```
1953 updates: 2
```

Hide

```
cat('\n')
```

Hide

```
cat('1969 updates:', sum(!(munros$y1953 == munros$y1969)) )
```

```
1969 updates: 0
```

Hide

```
cat('\n')
```

Hide

```
cat('1974 updates:', sum(!(munros$y1969 == munros$y1974)) )
```

```
1974 updates: 33
```

Hide

```
cat('\n')
```

Hide

```
cat('1981 updates:', sum(!(munros$y1974 == munros$y1981)) )
```

```
1981 updates: 102
```

Hide

```
cat('\n')
```

Hide

```
cat('1984 updates:', sum(!(munros$y1981 == munros$y1984)) )
```

```
1984 updates: 10
```

Hide

```
cat('\n')
```

Hide

```
cat('1990 updates:', sum(!(munros$y1984 == munros$y1990)) )
```

```
1990 updates: 0
```

Hide

```
cat('\n')
```

Hide

```
cat('1997 updates:', sum(!(munros$y1990 == munros$y1997)) )
```

```
1997 updates: 38
```

Hide

```
cat('\n')
```

Hide

```
cat('2021 updates:', sum(!(munros$y1997 == munros$y2021)) )
```

```
2021 updates: 10
```

The **1981 survey had the largest change in classification of Munro statues**, with 102 changes since the previous survey held in 1974. Twice, a survey yielded no updates in Munro class changes (1969 and 1990 compared to the previous of 1953 and 1984, respectively).

### 3) Which Munro is the most remote?

I will use the results of the 2021 survey to show recent statuses





Clearly, there is a munro located in the South East that is far from the rest. I can filter the data by comparing the coordinate values.

Hide

```
mask1 <- munros[munros$y2021 == 'Munro',] # Filter the Munros in 2021
mask2 <- mask1[mask1$xcoord < 160000, ] # Filter the xycoordinates
mask2[mask2$ycoord < 750000,]
```

Name	Height_m	Height_ft	xcoord	ycoord	y1891	y1921	y1933	y1953	y1969
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<chr>	<chr>	<chr>	<chr>
Ben More	966	3169	152576	733078	Munro	Munro	Munro	Munro	Munro

1 row | 1-10 of 16 columns

**Ben More** is the name of the distant Munro in the South East.

While I am sure there is a mathematical way to solve this, I feel that my answer is correct and can cover the solution.

# Finished

Looking back:

- It has been a while since I last used R. Almost 2 years! Thus, this took a while. I do believe I prefer Python and Matplotlib over R and ggplot. Or maybe it's because I just got so used to Python. Either way, was a good refresher.

- Found out that R dataframes can not have column names made up of only integers. Which makes sense in hindsight, but I really got tripped up for a while during the analysis portion, before going back and reformatting the names.
- A little annoyed on the 2 legends in my scatter plot of the Munro locations. If I plotted in Matplotlib, I feel like I could combine the two legends. But oh well. I am less experienced with R.