**Pls explain and arrange in your own words. Don’t copy these directly cuz they are incomplete. Keywords are given in case you don’t remember and feel free to ask me anything**

**Why use RF:**

**Strengths that you google online (use voting to overcome overfitting, etc)**

**Better than Logistics cuz we tried**

**Process:**

1. The previous input dataset is useless, pls use the new one here. Pls expand variable’s names to make them easily understandable.

For the summary in the plots. First we should explain how we selected them,:

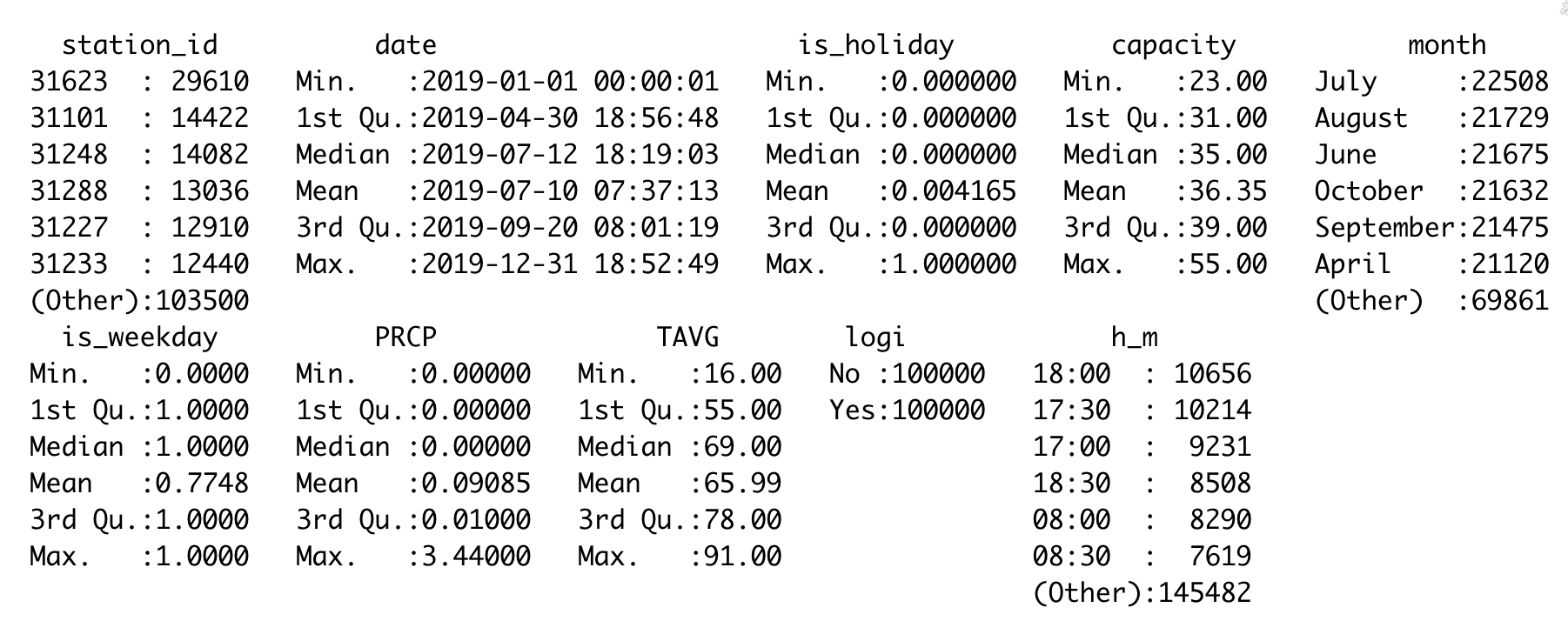
**Select 20 stations**

**stratified sampling, each 50%**

**Use 15% as threshold to create response**

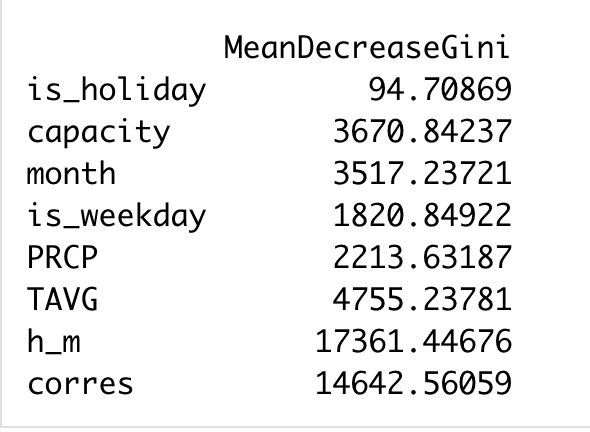
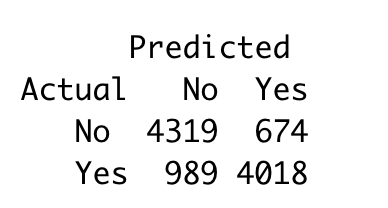
Then make a table remember align and comma:

**(delete “capacity” for below output! )**



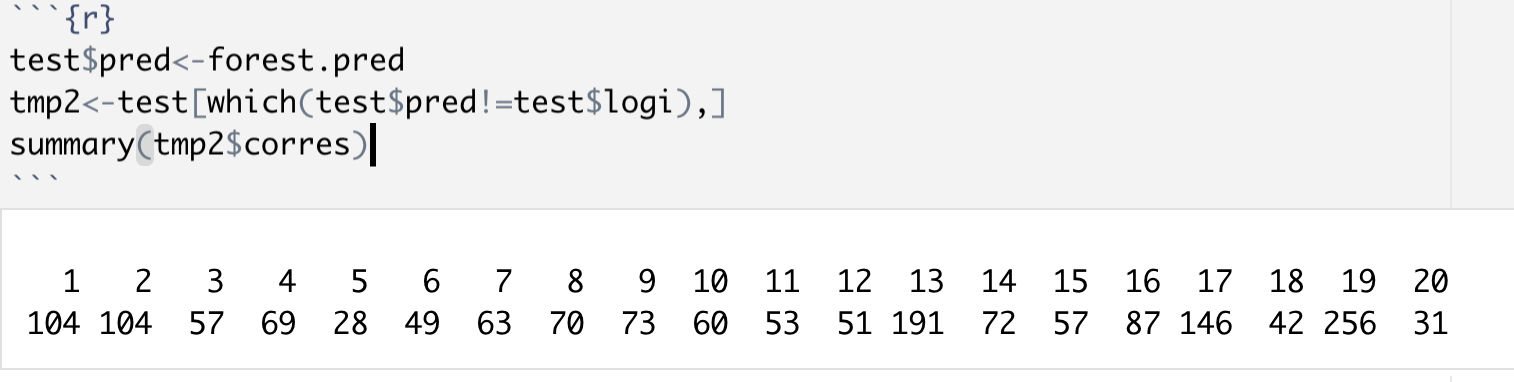
For this one select some statistics to include in the table.

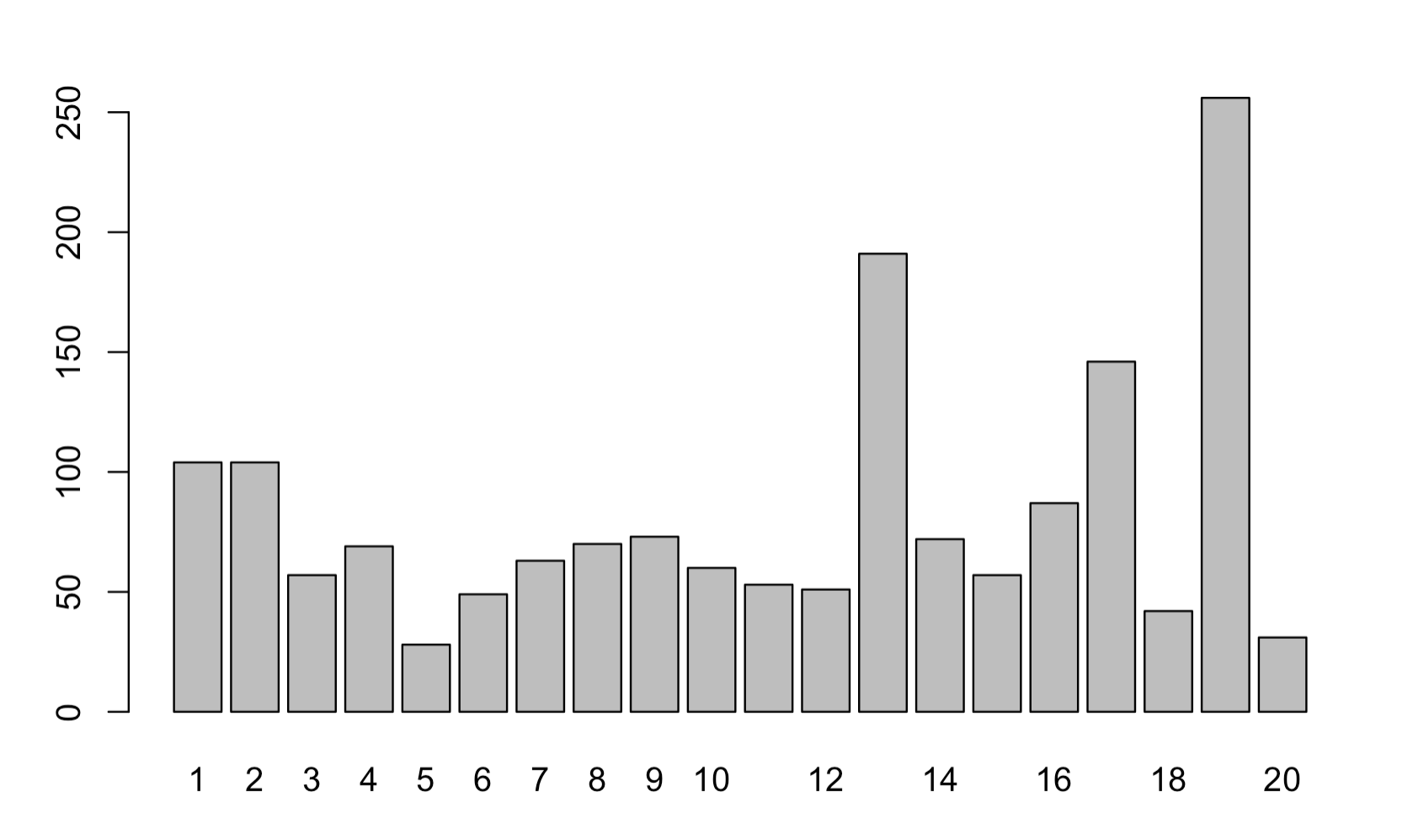
Then mention we split the dataset, it’s better to make a flowchart here.

1. Make a table about importance (no values just importane) use the requirements above for these (use descending order) 
2. Make a table 

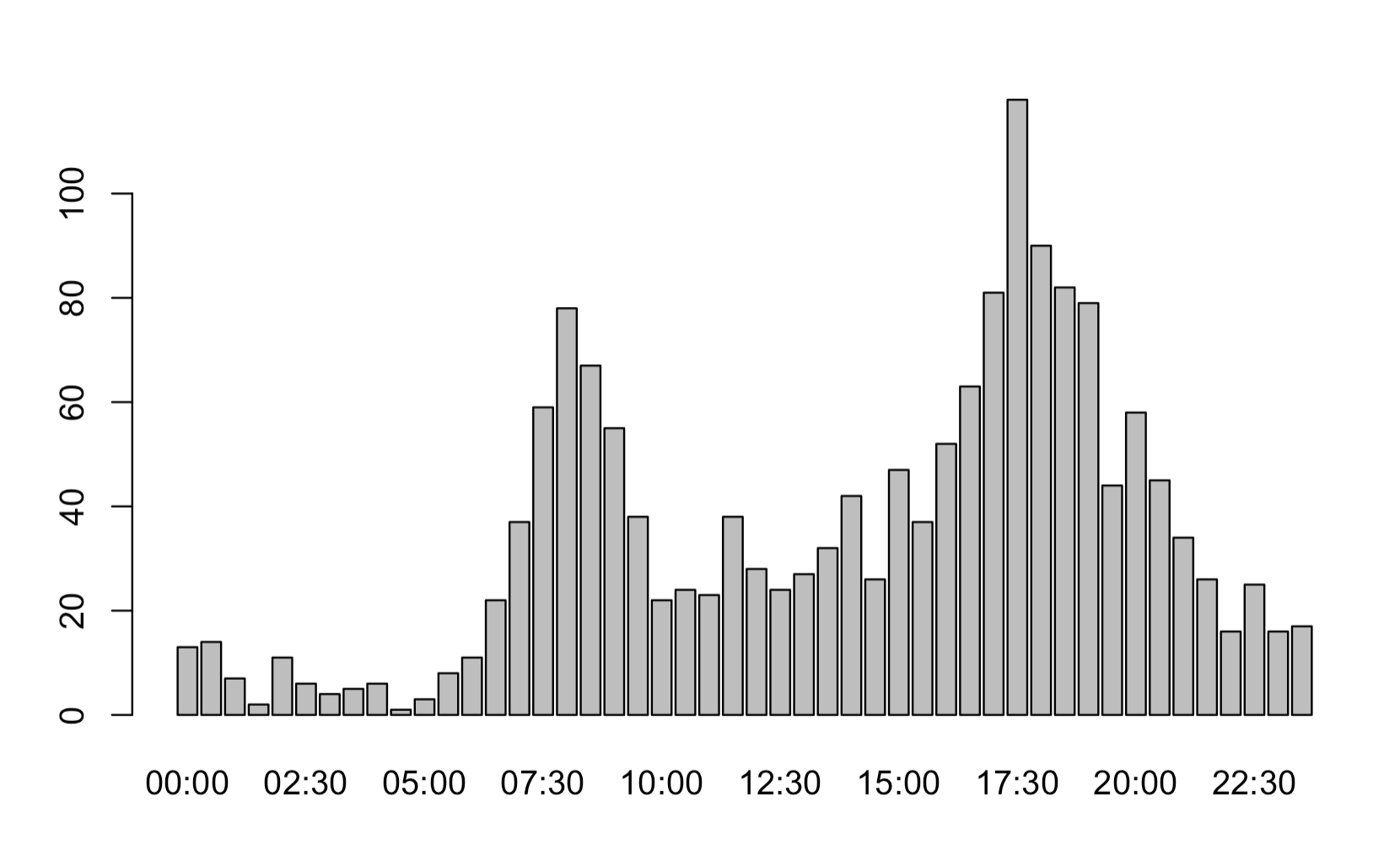
Accuracy 83.37%

1. EDA on clustered wrong



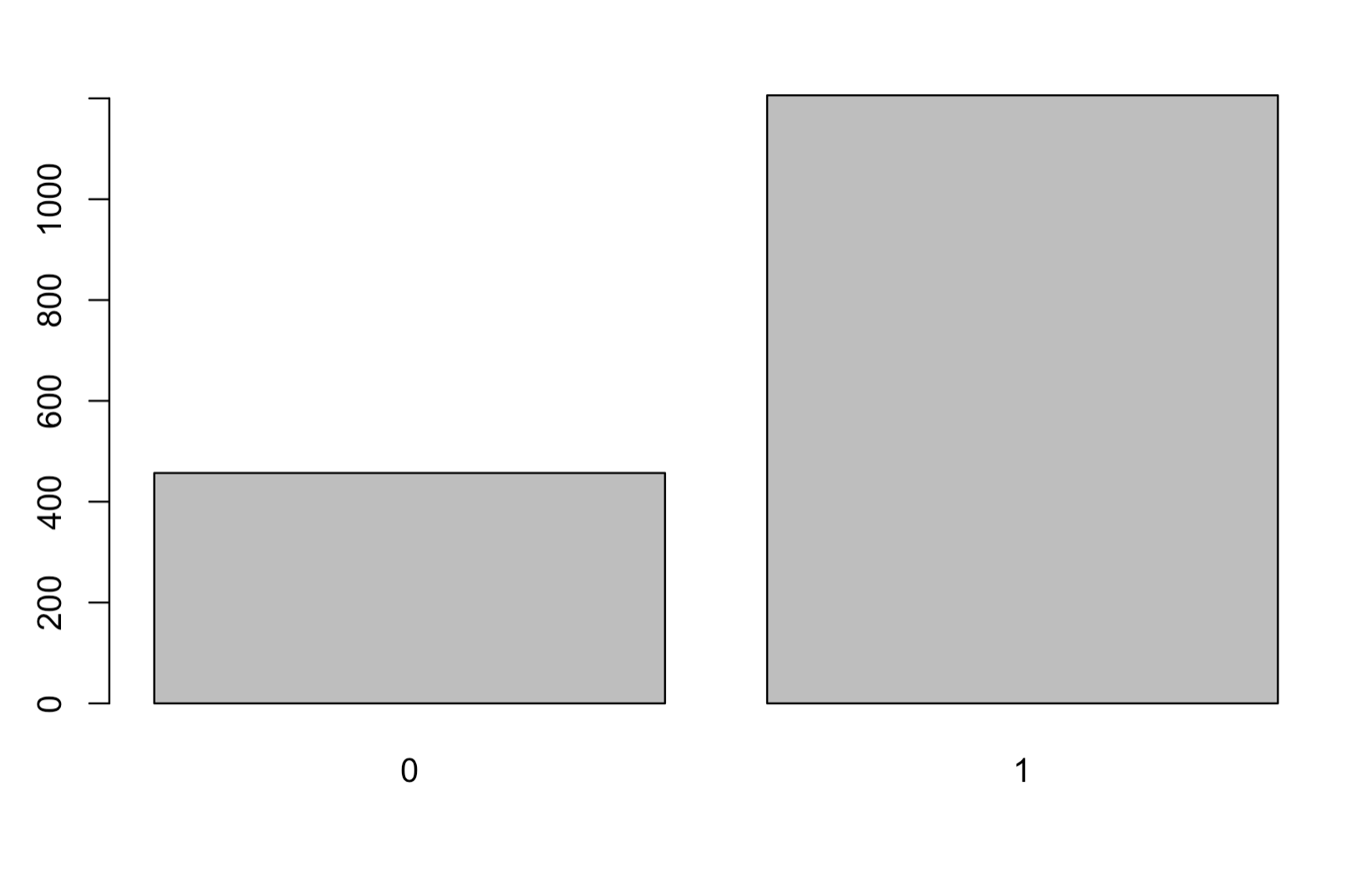


For stations\_id

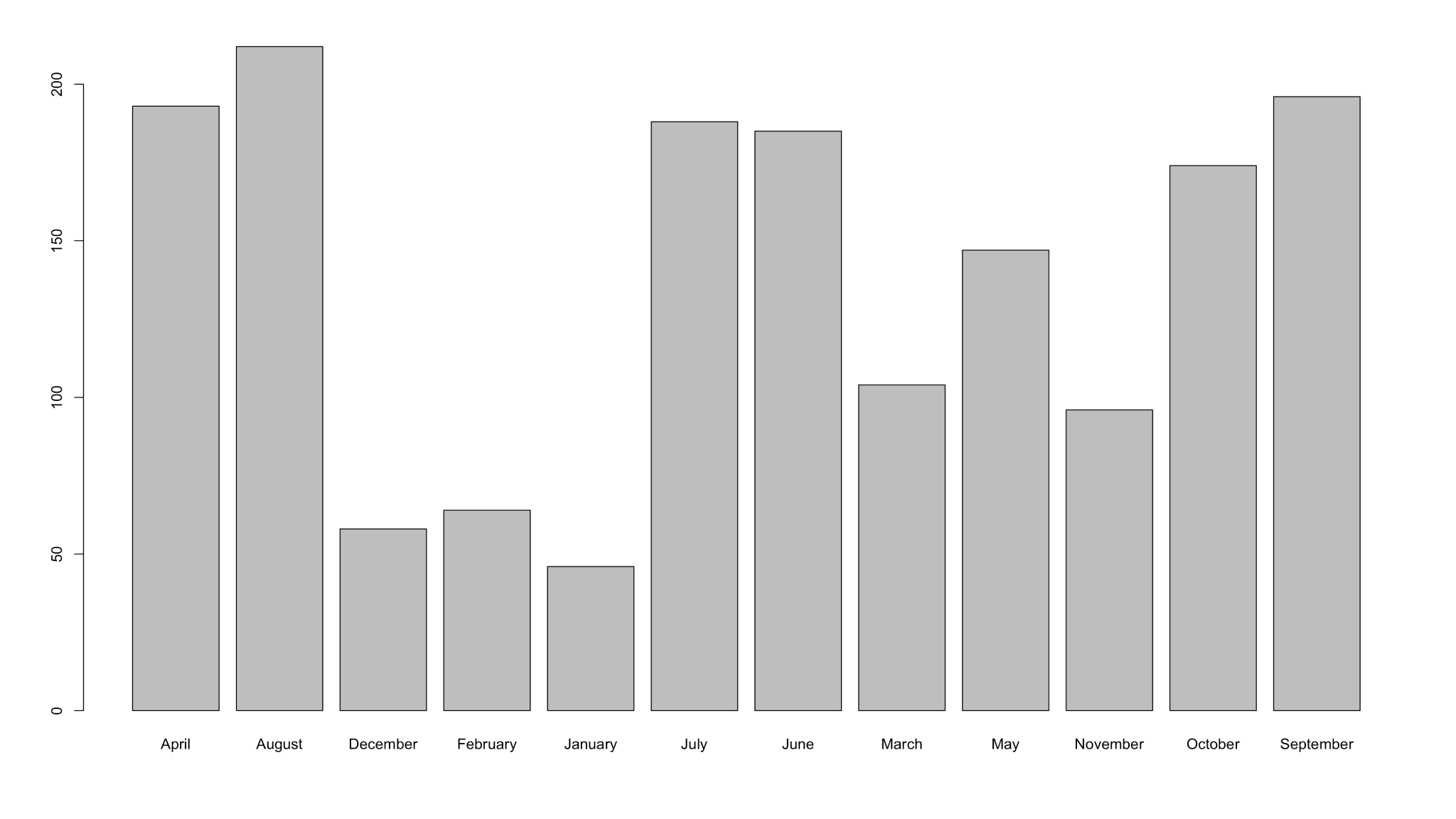


For h\_m’

Is\_weekday



month



Is\_holiday

