Lecture 2: Exercise & Answers

To become a data ninja requires you one to be a master of basic extraction of data.

The Scenario

Organizations always need some time to warm up to data. Often times, simple statistical reports are a way to build trust. Given the following data frame containing financials from organizational operations, write the appropriate expression to extract the right information.

Note that `revenue` is total revenue in millions, `rev_target` is the target revenue for a given month, `elec_bill` is the electricity bill in thousands, and `social_media` is the number of likes. All data are monthly numerical values.

```
##    month quarter revenue rev_target elec_bill social_media
## 1      1       1     266        284       128           80
## 2      2       1     289        291       141          257
## 3      3       1     302        214       145           14
## 4      4       2     138        124       207          133
## 5      5       2     121        104       194          240
## 6      6       2     268        172       154           37
## 7      7       3     299        266       213          168
## 8      8       3     246        299       191           62
## 9      9       3     217        241       176           38
## 10    10       4     133        146       187          226
## 11    11       4     171        100       160          269
## 12    12       4     174        146       153          112
```

The Questions

Assuming the data is in data frame format, write commands to answer the following questions:

Example:

How many months of data do you have?

```
nrow(df$month)
```

```
## NULL
```

1. What was the total revenue for the year?
2. What was the most number of tweets?
3. How much was the total elec bill for Q2 and Q3?
4. What were the last 5 monthly electricity bill amounts and revenues?
5. In which month did we hit our lowest number of likes?
6. In which quarter did we hit our highest revenue?
7. In how many months did we hit our revenue targets?
8. In which months did we hit our target?
9. In months where we missed the target, what was the total lost revenue?
10. Provide the electric bills for the first month of each quarter.

Answers

1. What was the total revenue for the year?

```
  sum(df[,3])
  sum(df$revenue)
```

2. What was the most number of tweets?

```
  max(df[,6])
  max(df$social_media)
```

3. How much was the total elec bill for Q2 and Q3?

```
  sum(df[4:9, 5])
```

```
## [1] 1135
```

```
  sum(df[df$quarter == 2 | df$quarter == 3, "elec_bill"])
```

```
## [1] 1135
```

4. What were the last 5 monthly electricity bill amounts and revenues?

```
#Explicit
  df[7:12, c(3,5)]
```

```
##    revenue elec_bill
## 7      299       213
## 8      246       191
## 9      217       176
## 10     133       187
## 11     171       160
## 12     174       153
```

```
  df[df$month > 7, c("revenue", "elec_bill")]
```

```
##    revenue elec_bill
## 8      246       191
## 9      217       176
## 10     133       187
## 11     171       160
## 12     174       153
```

```
#Based on relative position in data
  row <- nrow(df)
  df[(row-4):row, c(3,5)]
```

```
##    revenue elec_bill
## 8      246       191
## 9      217       176
## 10     133       187
## 11     171       160
## 12     174       153
```

5. In which month did we hit our lowest number of likes?

```
  df[min(df$social_media) == df$social_media, 1]
```

6. In which quarter did we hit our highest revenue?

```
df[max(df$revenue) == df$revenue, 2]
```

7. In how many months did we hit our revenue targets?

```
nrow(df[df$revenue > df$rev_target,])
```

8. In which months did we hit our target?

```
df[df$revenue > df$rev_target, "month"]
```

9. In months where we missed the target, what was the total lost revenue?

```
offtimes <- df$revenue < df$rev_target
sum(df$rev_target[offtimes] - df$revenue[offtimes])
```

10. Provide the electric bills for the first month of each quarter.

```
df[df$month %% 3== 1, 5]
```