sdss

Victoria

6/11/2021

Documentation

Setting up my environment

```
library(tidyverse)
```

Queries

Here, I want to explore and analyze an aspect of the Sloan Digital Sky Survey DATA RELEASE 7. This is not the latest data release from the SDSS, however due to accessibilty, it is enough to suit my purpose.
I obtained the data set through the BigQuery database, and dabbled with some SQL queries.

This is a rather large dataset so I wanted to minimize it to items I could manage and had some knowledge with which I wanted to explore.

Note

I was having a hard time connecting R studio to Bigquery, so I had to continue with downloading the data set I filtered out with the following SQL queries. This is something I aim to return to and correct at a later date. For now, I will omit the ““‘” at the beginning and end of the SQL queries.

{sql connection=con}
SELECT
NAME,
z
FROM
‘big-public-data.blackhole_database.sdss_dr7’

This presented an issue because I was still seeing 7000+ items, and I truly just wanted information on a select few items: quasars and galaxies.
So, I took to the SDSS website and sought out any information that could help me narrow down the data. Luckily, I found a table that mapped the objects I was looking for that were scanned by PLATES 673-714. Luckily, this was an attribute on the dataset so I wrote out another quick query:

{sql connection=} SELECT
NAME,
z
FROM
’big-public-data.blackhole_database.sdss_dr7
WHERE
PLATES BETWEEN 673 AND 714

One more step to make it seem a little less chaotic: descend by z!

{sql connection=}
SELECT

NAME,
z
FROM
'big-public-data.blackhole_database.sdss_dr7
WHERE
PLATES BETWEEN 673 AND 714
ORDER BY
z

This was finally a dataset I felt I could work with comfortably and decided it was ready for export and analysis through R.

Some Statistical Analysis

I wanted to get a feel for the values in my data. Sure the names were long, however I knew I was working with galaxies and quasars which I have knowledge of, so I wanted to focus on the z column (z known as redshifts in astronomy).

```
sdss_df <- read.csv('sdss.csv')
```

Next, I wanted to find the minimum and maximum redshifts:

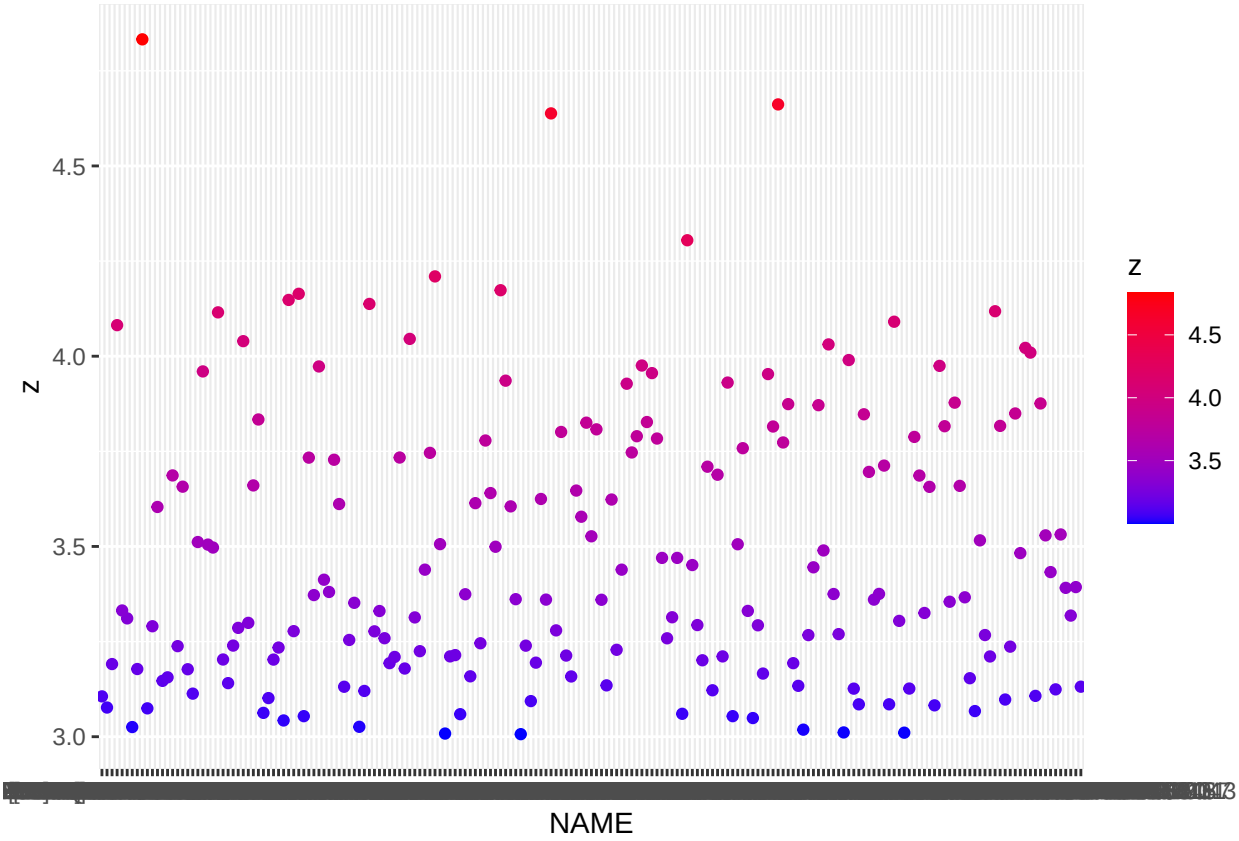
```
min(sdss_df['z'])
```

```
## [1] 3.0067
```

```
max(sdss_df['z'])
```

```
## [1] 4.8327
```

```
ggplot(data = sdss_df) +
  geom_point(mapping= aes(x=NAME, y=z, color=z)) +
  scale_color_gradient(low="blue", high="red")
```



More analysis and inference done within presentation.