607 Week 2 – Snowbird SQL

Ethan Haley

2/13/2021

Snowbird Chairlifts and Trails

Snowbird is one of the world's top ski mountains, located up Little Cottonwood Canyon from the Salt Lake Basin in Utah, in a spot where impressive amounts of perfectly skiable snow pile up every year. Here's a look at the trail map for the back and front sides of the mountain:

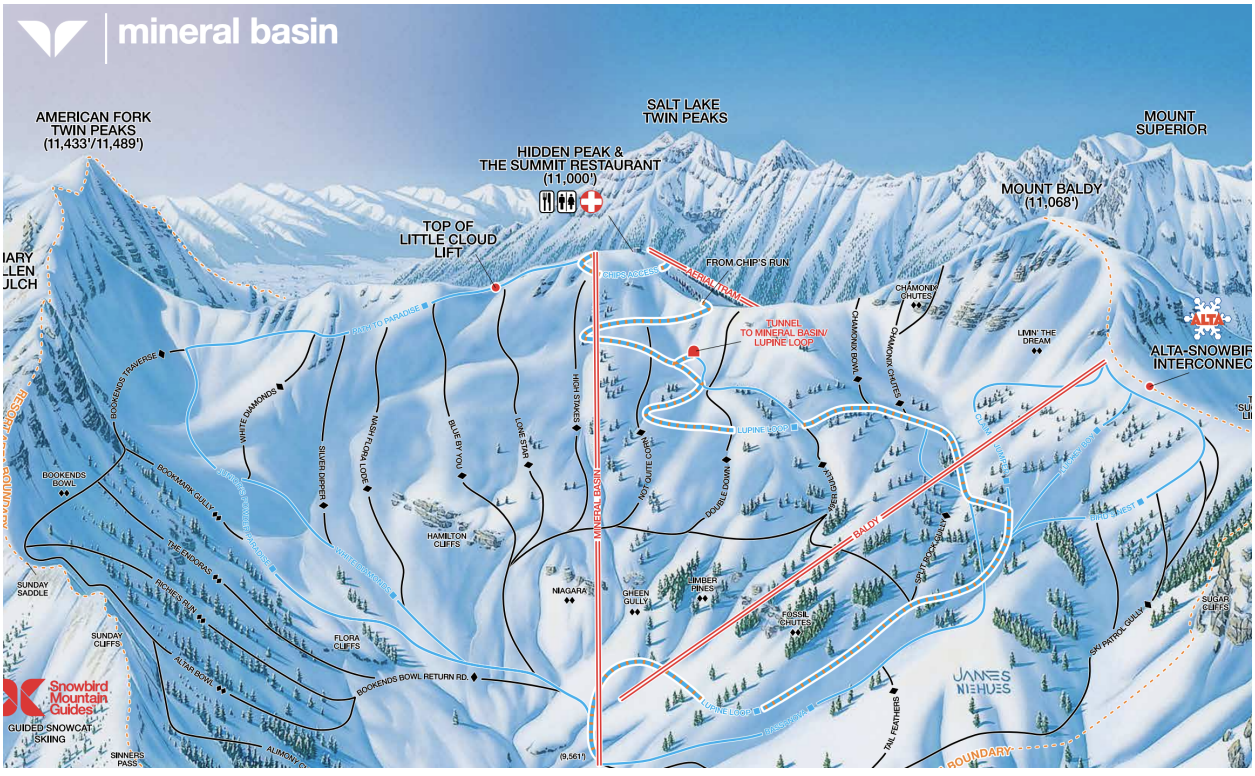


Figure 1: Back Side of Snowbird

Figure 2: Front Side of Snowbird

There's a lot of information in these standard trail maps, and snow sports enthusiasts could spend all day looking at them. But even a frequent Snowbird visitor, perhaps relegated to building and querying databases about the mountain he's unable to get to during the pandemic, can tell you that there's much to learn about the layout of the mountain from an abstracted, data-driven look at what you see on the map.

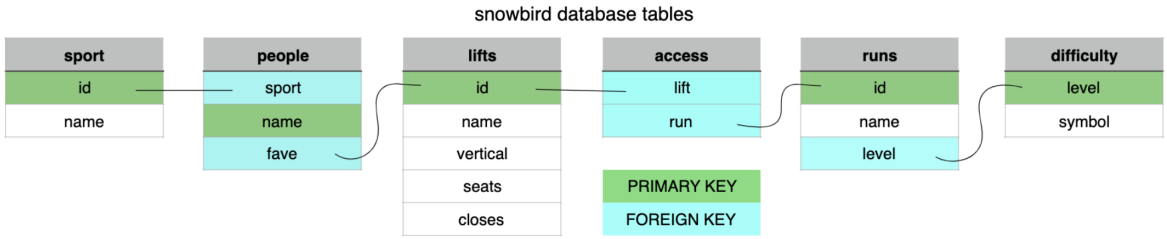
A few important and common questions that I'll at least provide a structure for answering in this brief analysis are:

- "Based on my abilities, which chairlift should I take?"
- "My group is going to split up for a couple of hours. Where should we meet up later?"
- "I'm a skier, not a snowboarder. Is there a part of the mountain I might prefer?"
- "Which lift gives you the most choices of runs to take?"
- "It's my last day here. How can I get as many good runs in as possible before lifts close?"

Here's where the data to answer these questions come from, and how they're organized:

I used PostgreSQL in pgAdmin 4 to build a database of six normalized tables, which amounts to a "toy example", but which offers a great opportunity to learn about database architecture and also provides a solid scaffolding for a more thorough analysis and representation of the map/topology of the mountain. The data all come from the publicly available trail map at https://snowbird.com, but knowledge of which `lifts` access which `runs`, and vice versa, isn't always completely obvious from a glance at the map. That's where I used my domain knowledge in the `access` table of the database, which models a many-to-many relationship between chairlifts and the runs they access. Rather than pointing this out later, when you'll have forgotten this background, the most important next step to this project is to add an analogous table which represents which *runs* access which *lifts*. The value there is in making a complete map of how you can get from wherever you are to wherever you want to be, which is a constant struggle on a huge and often unfamiliar mountain.

The relationship between the six tables is diagrammed here:



Although I did include most of the lifts, I only included about a quarter of all runs, specifically the ones that my family and our friends have enjoyed on group trips over the years, providing a nice sample range of skill levels and terrain. The `people` table lists just 8 of us so far, along with our favorite chairlift and preferred sport (snowboarding/skiing), and will be very easy to build to a size where it's actually pretty useful as a source of advice for other people. In case it's not obvious yet, skiers love talking about their favorite anything!

For example, let's take a look at the first question above, "Based on my abilities, which chairlift should I take?"

```
liftLevels <- data.frame(read_csv("SQL_csvs/lift-run-combos.csv"))
head(liftLevels)
```

```
##            lift vertical seats            run level
## 1   aerial tram     2900   125     chip's run     2
## 2      peruvian     2572     4     chip's run     2
## 3   aerial tram     2900   125 upper primrose     3
## 4  little cloud     1304     4     old ladies     4
## 5 mineral basin     1429     4     old ladies     4
## 6   aerial tram     2900   125     old ladies     4
```

The last column, `level`, contains the info that every skier uses to describe herself, the standard "green", "blue", "black", "double black" scale used in most of the U.S.

```
difficulty <- data.frame(read_csv("SQL_csvs/difficulty.csv"))
head(difficulty)
```

```
##   level               symbol
## 1     1         green circle
## 2     2          blue square
## 3     3        black diamond
## 4     4 double black diamond
```

Now if someone said "I like to ski blues mostly", we could look up which lifts serviced such terrain with the highest ratio. The reason we shouldn't instead just look up which lift accesses the *most* blues is that lifts that go to the very top of the mountain tend to access the most runs, but they often force a skier to begin with a more difficult run, before leveling out into the blue area.

```
blues <- liftLevels %>% filter(level == 2)
bluesTable <- table(blues$lift)
liftTable <- table(liftLevels$lift)
bluesTable
```

```
##
##   aerial tram         gad 2       gadzoom  little cloud mineral basin
##             6             5             3             4             4
##      peruvian       wilbere
##             2             1
```

```
liftTable
```

```
##
##   aerial tram     chickadee         gad 2       gadzoom  little cloud
##            24             1            15             9            17
## mineral basin      peruvian       wilbere
##            19             8             3
```

TODO – 1) learn R so that I can plot the most basic of concepts, like how to get the proportion of blues for each lift, from the table. Or how to group_by/aggregate. And how to do the equivalent of a join, but in R.

"My group is going to split up for a couple of hours. Where should we meet up later?" A good approach to answering this is to calculate the runs that are reachable from the most lifts:

```
accessibleRuns <- data.frame(read_csv("SQL_csvs/waysToRuns.csv"))
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   run = col_character(),
##   ways_there = col_double(),
##   symbol = col_character()
## )
```

```
head(accessibleRuns, n=10)
```

```
##                    run ways_there               symbol
## 1          bass highway          7         green circle
## 2         wilbere ridge          6          blue square
## 3        harper's ferry          6        black diamond
## 4          fluffy bunny          5        black diamond
## 5             gad gully          5        black diamond
## 6        upper big emma          5          blue square
## 7  middle bassackwards          5          blue square
## 8    lower/middle emma          5         green circle
## 9          jones avenue          5 double black diamond
## 10              niagara          4 double black diamond
```

"I'm a skier, not a snowboarder. Is there a part of the mountain I might prefer?" Using the favorite lifts and sport columns of people.csv:

```
people <- data.frame(read_csv("SQL_csvs/people.csv"))
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   name = col_character(),
##   favelift = col_character(),
##   sport = col_character()
## )
```

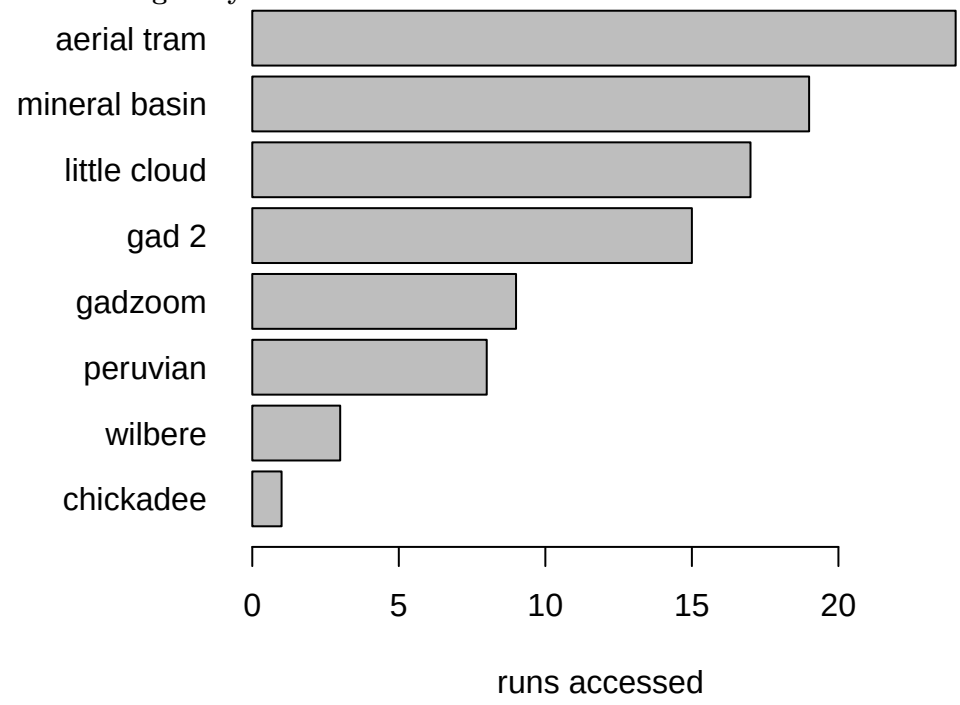
```
people
```

```
##      name favelift sport
## 1   ethan        3     1
## 2     kim        6     1
## 3   reese        4     1
## 4     viv        2     2
## 5   steve        2     2
## 6   alice        2     2
## 7   tommy        1     2
## 8 maureen     NULL  NULL
```

The skiers in this sample prefer lift 2, which is Peruvian.

```
par(mar = c(4, 10, 4, 4))
barplot(sort(table(liftLevels$lift)), horiz = T, las=1, xlab = "runs accessed")
```

"Which lift gives you the most choices of runs to take?"



"It's my last day here. How can I get as many good runs in as possible before lifts close?"
Order the lifts by closing times to see which ones you need to do first:

```
lifts <- data.frame(read_csv("SQL_csvs/lifts.csv"))
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   id = col_double(),
##   lift = col_character(),
##   vert_feet = col_double(),
##   seats = col_double(),
##   last_chair = col_time(format = "")
## )
```

```
lifts
```

```
##   id          lift vert_feet seats last_chair
## 1  3 mineral basin      1429     4   15:30:00
## 2  8   aerial tram      2900   125   15:45:00
## 3  5  little cloud      1304     4   15:45:00
## 4  4         gad 2      1242     4   16:00:00
## 5  7       gadzoom      1827     6   16:00:00
```

```
## 6  2      peruvian      2572     4   16:00:00
## 7  6       wilbere       668     2   16:30:00
## 8  1     chickadee       149     2   17:00:00
```

Figure 3: Just Before the Rope Drops at Mineral Basin