

GR5702

Exploratory Data Analysis and Visualization

Prof. Joyce Robbins

March 2, 2017

Patterns of missing data

MCAR = missing completely at random (no pattern)

MAR = missing at random (some pattern but it doesn't matter, can be accounted for)

MNAR = missing not at random (RED FLAG)

Dealing with Missing Values

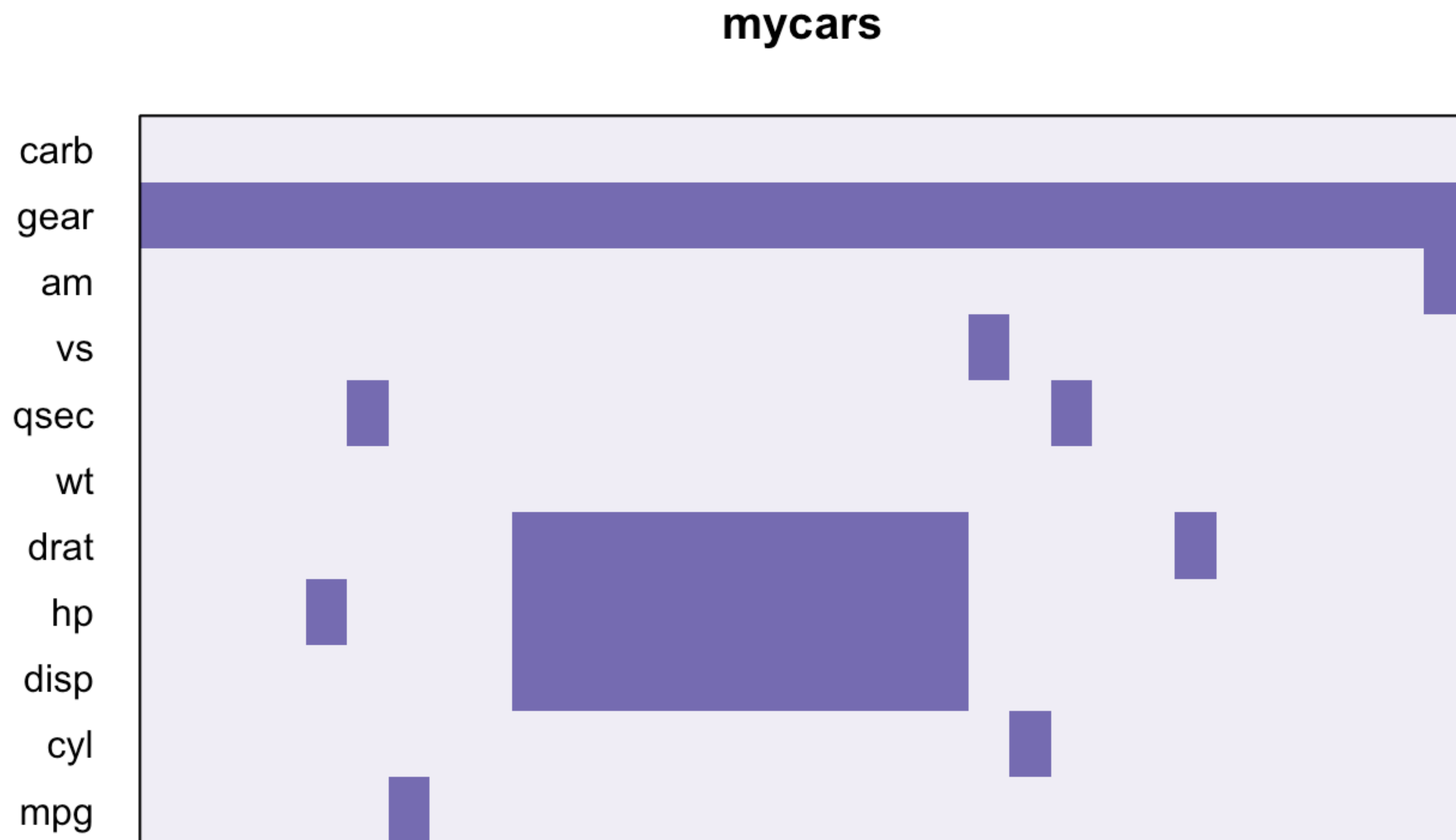
Step 1

Make sure all missing values are marked "NA"

```
library(readr)
```

```
df <- read_csv("...", na = c("", "NULL"))
```

variation of missing.pattern.plot (formerly in "mi" package)

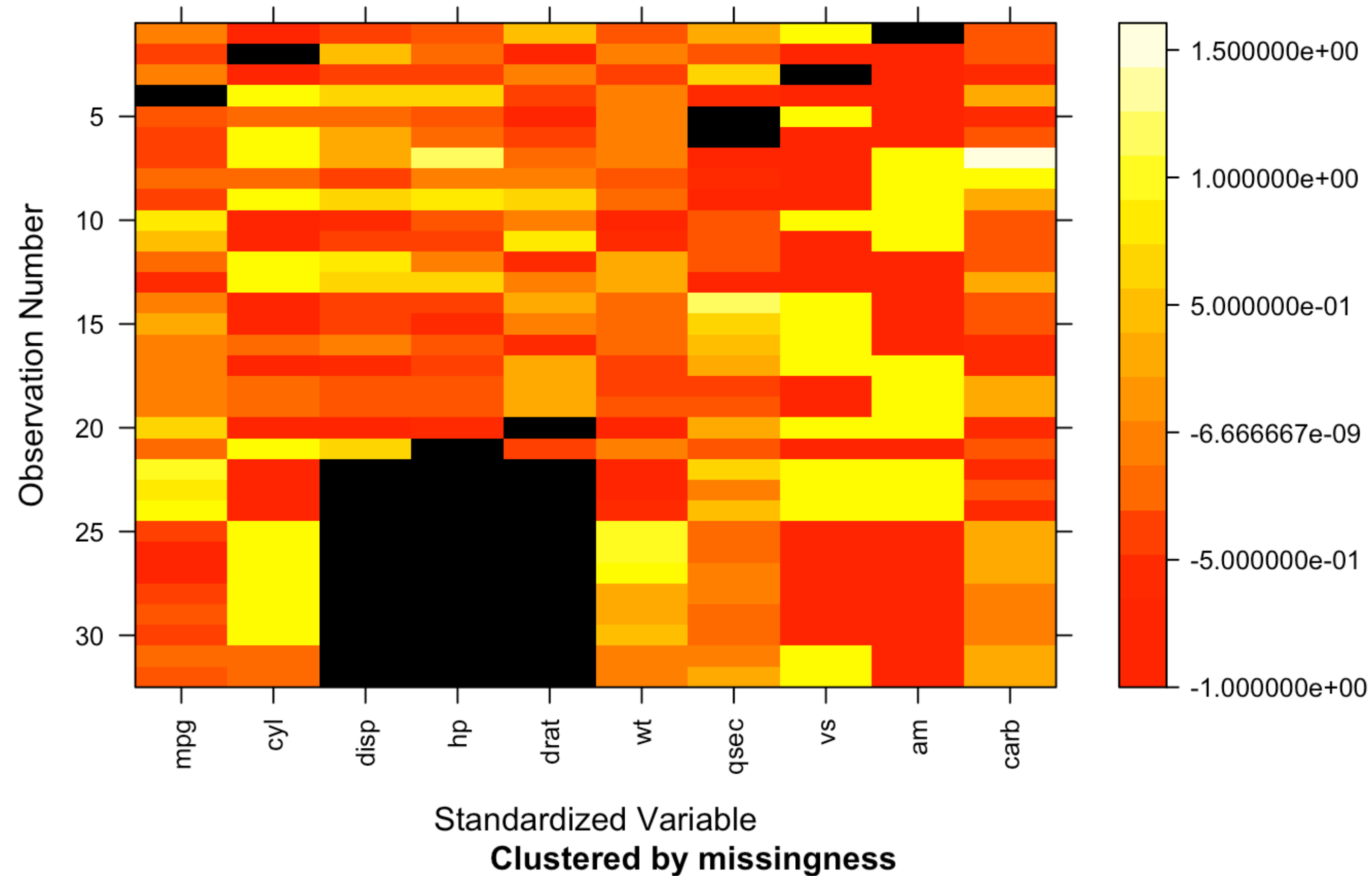


mycars



```
image(missing_data.frame())
```

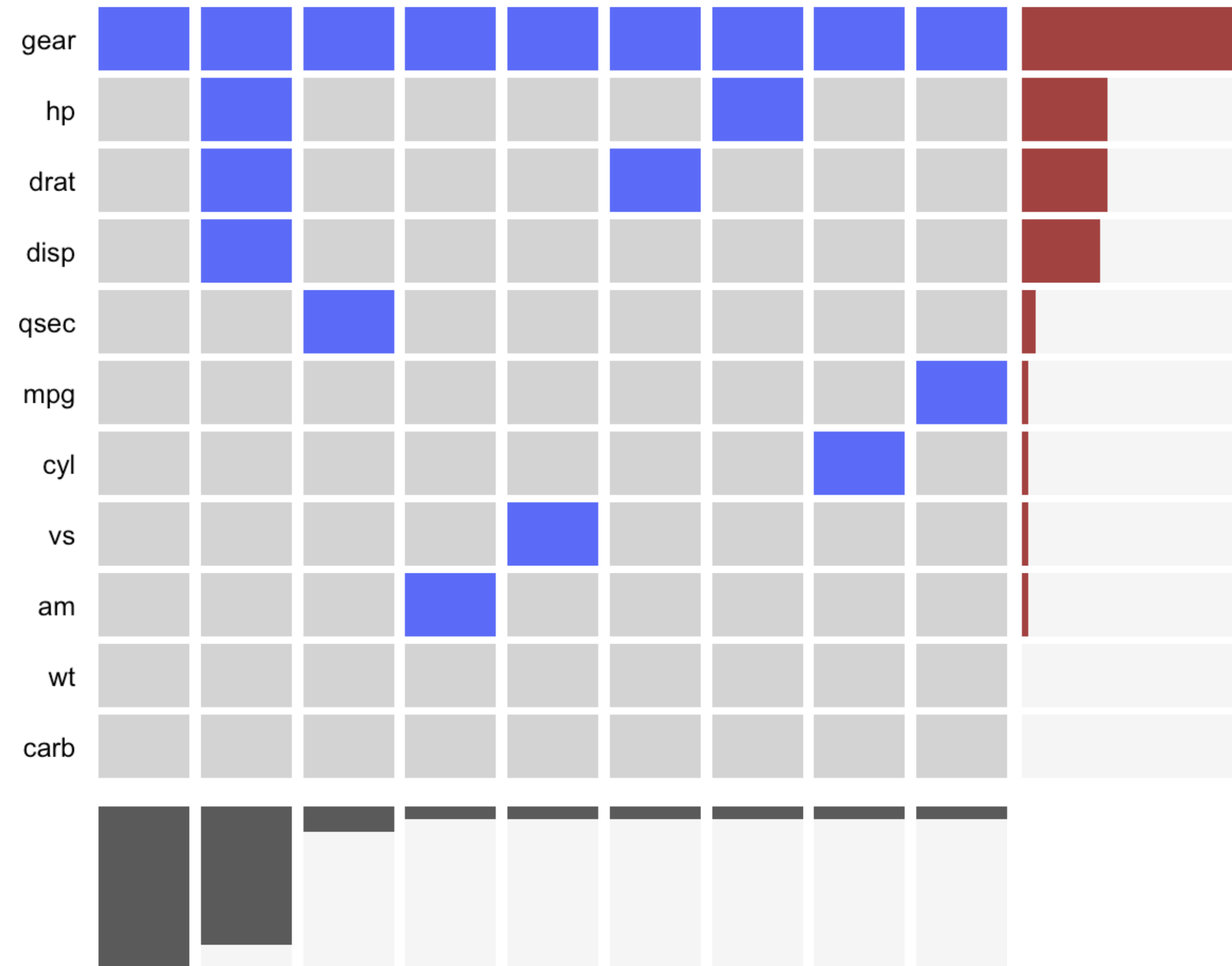
Dark represents missing data



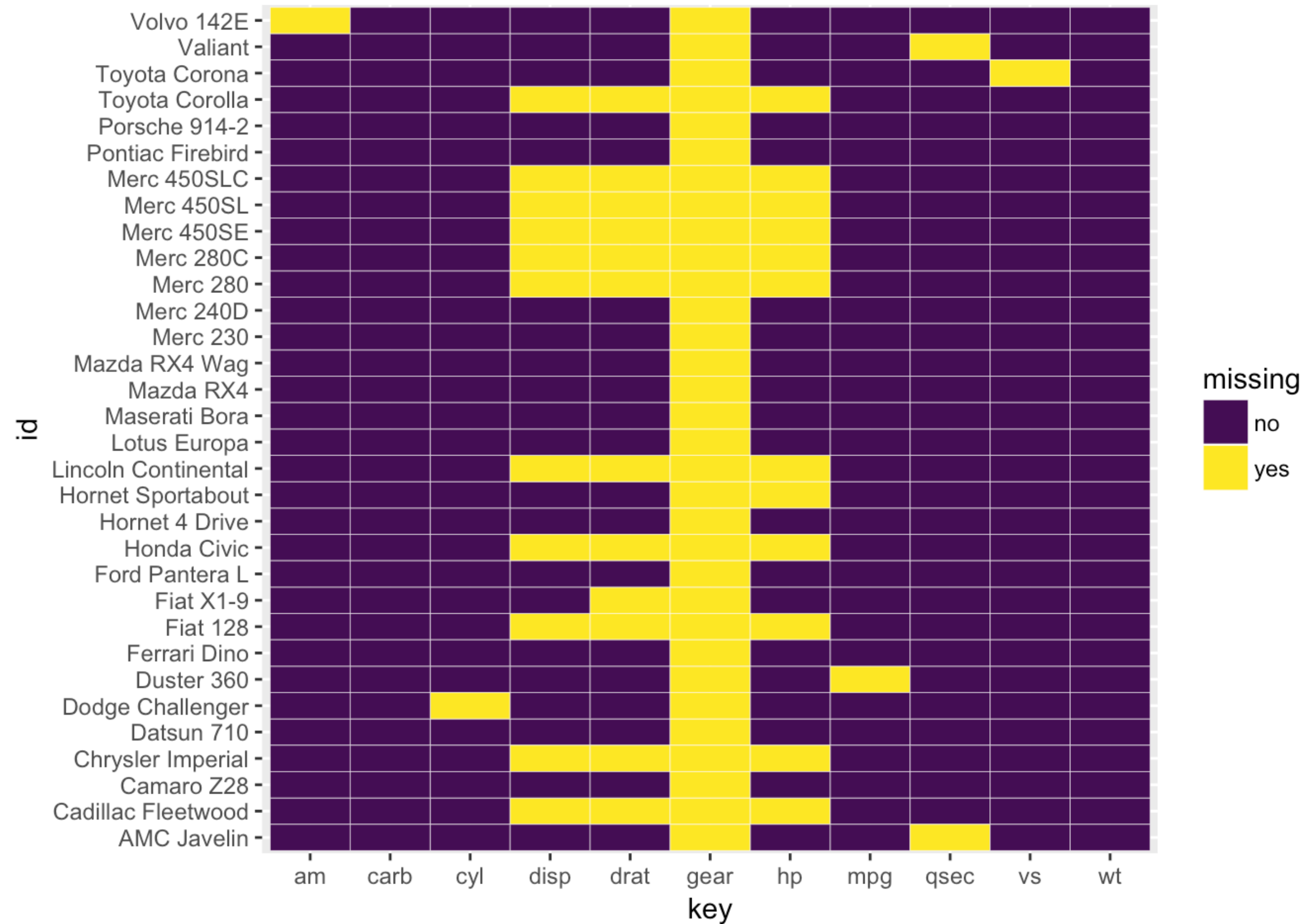
package:
mi

visna()

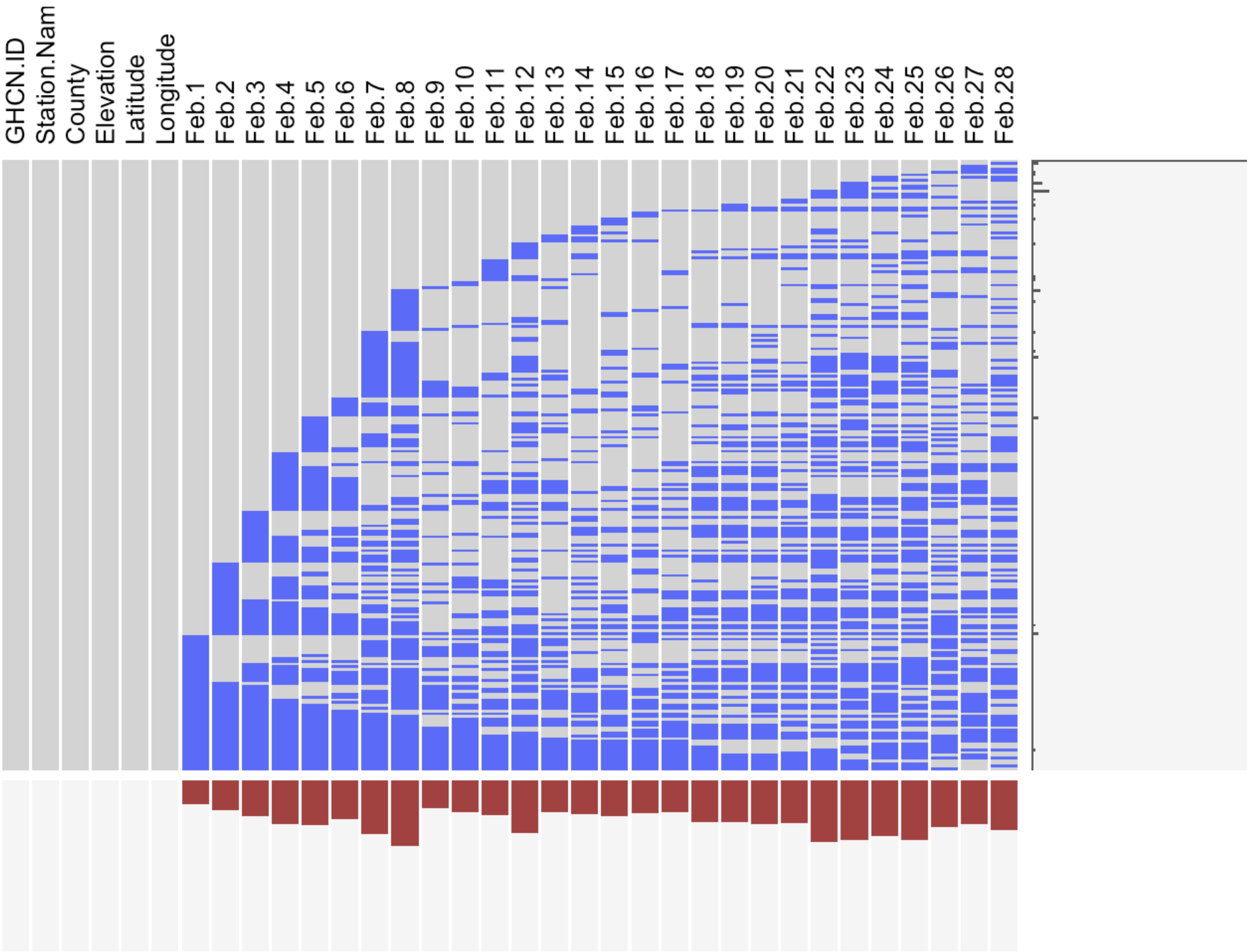
package:
extracat



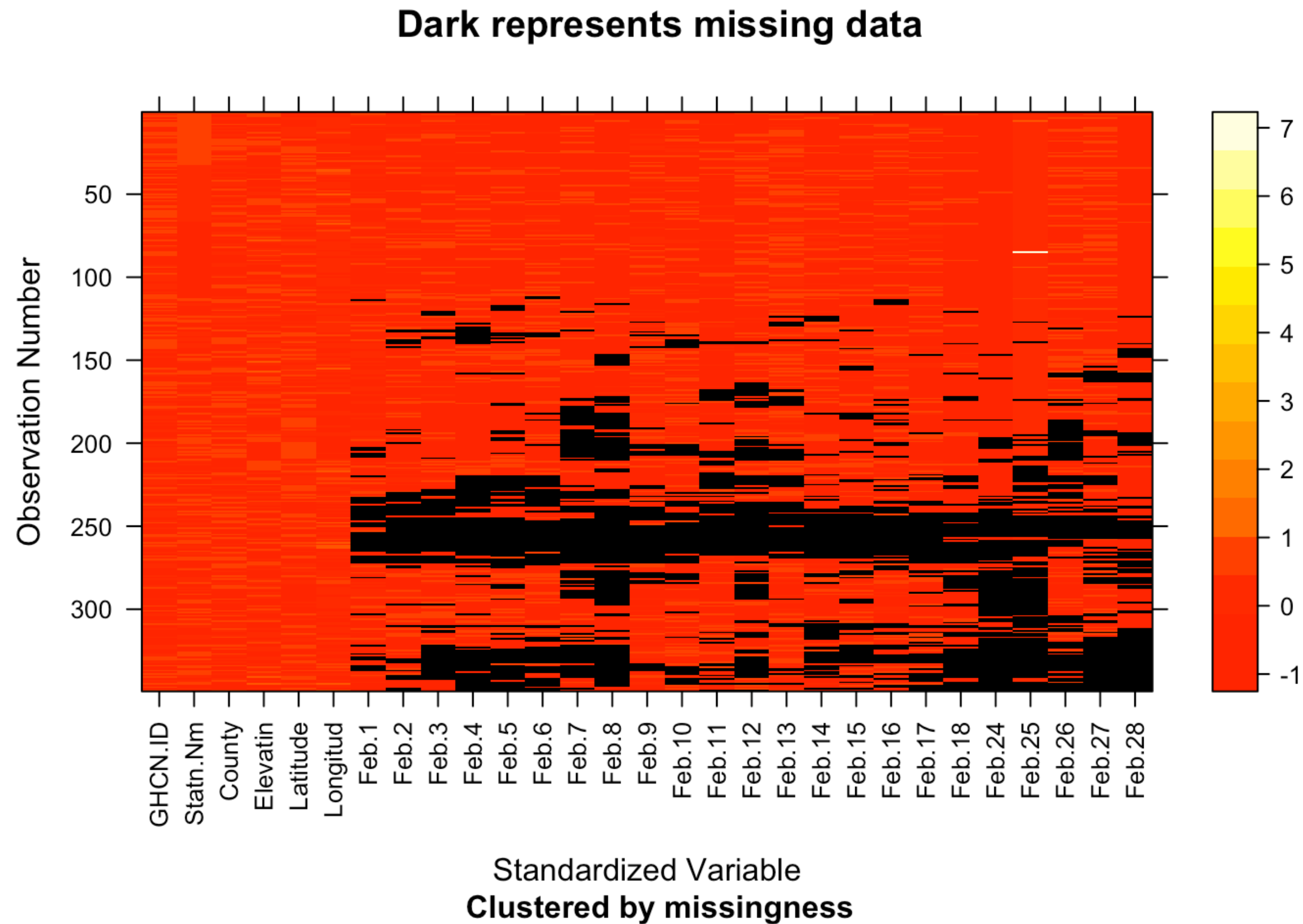
geom_tile()



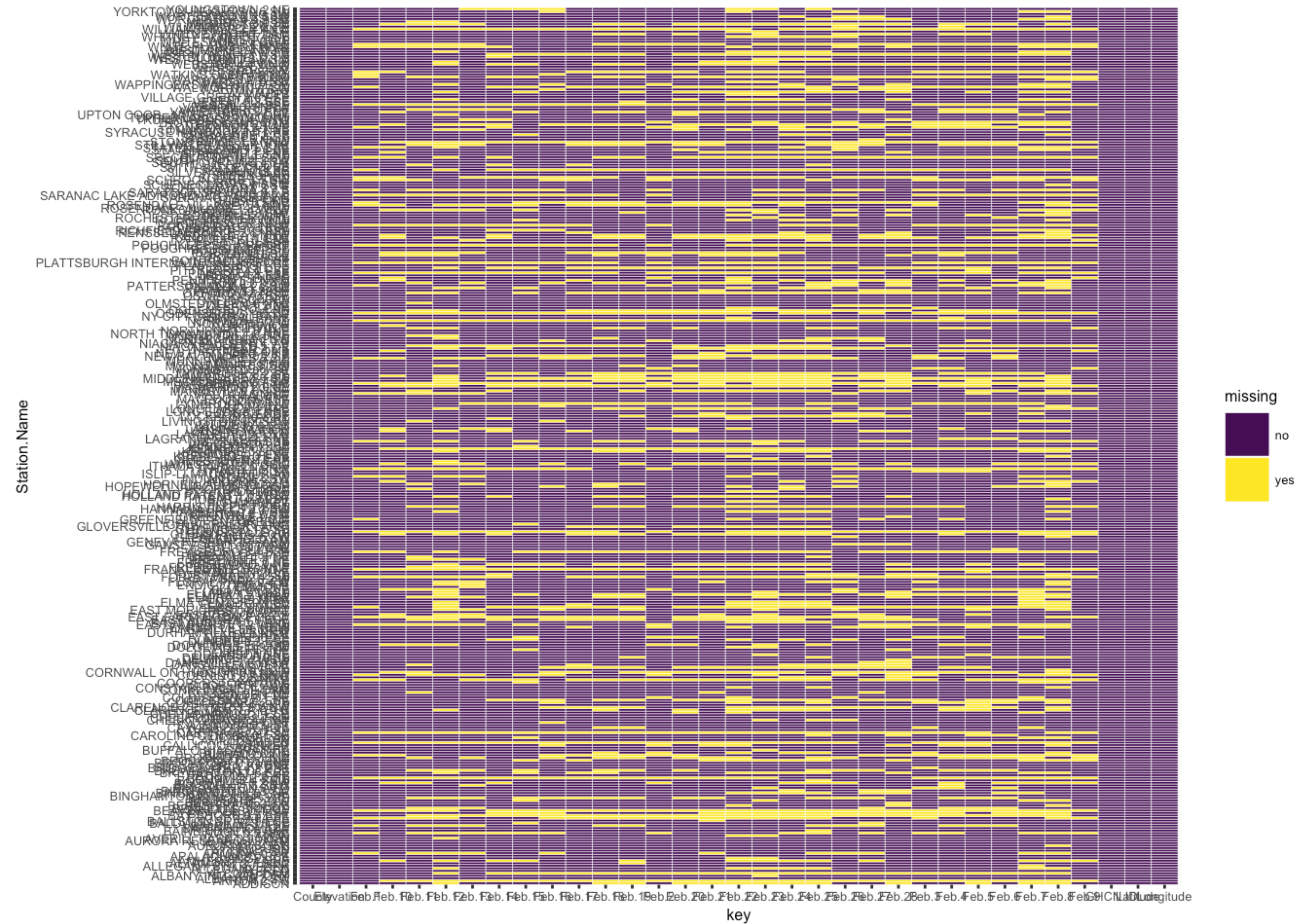
visna()



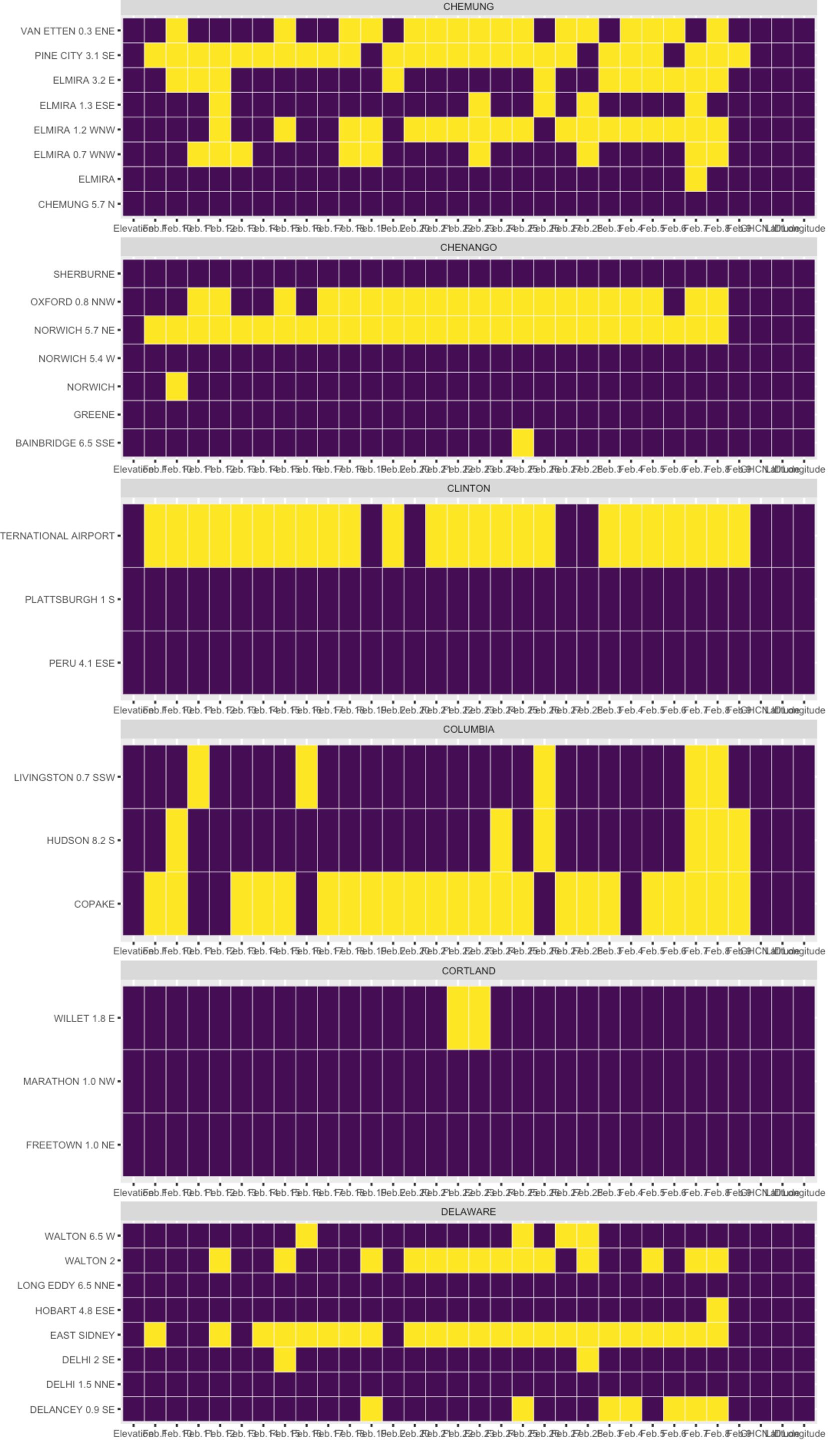
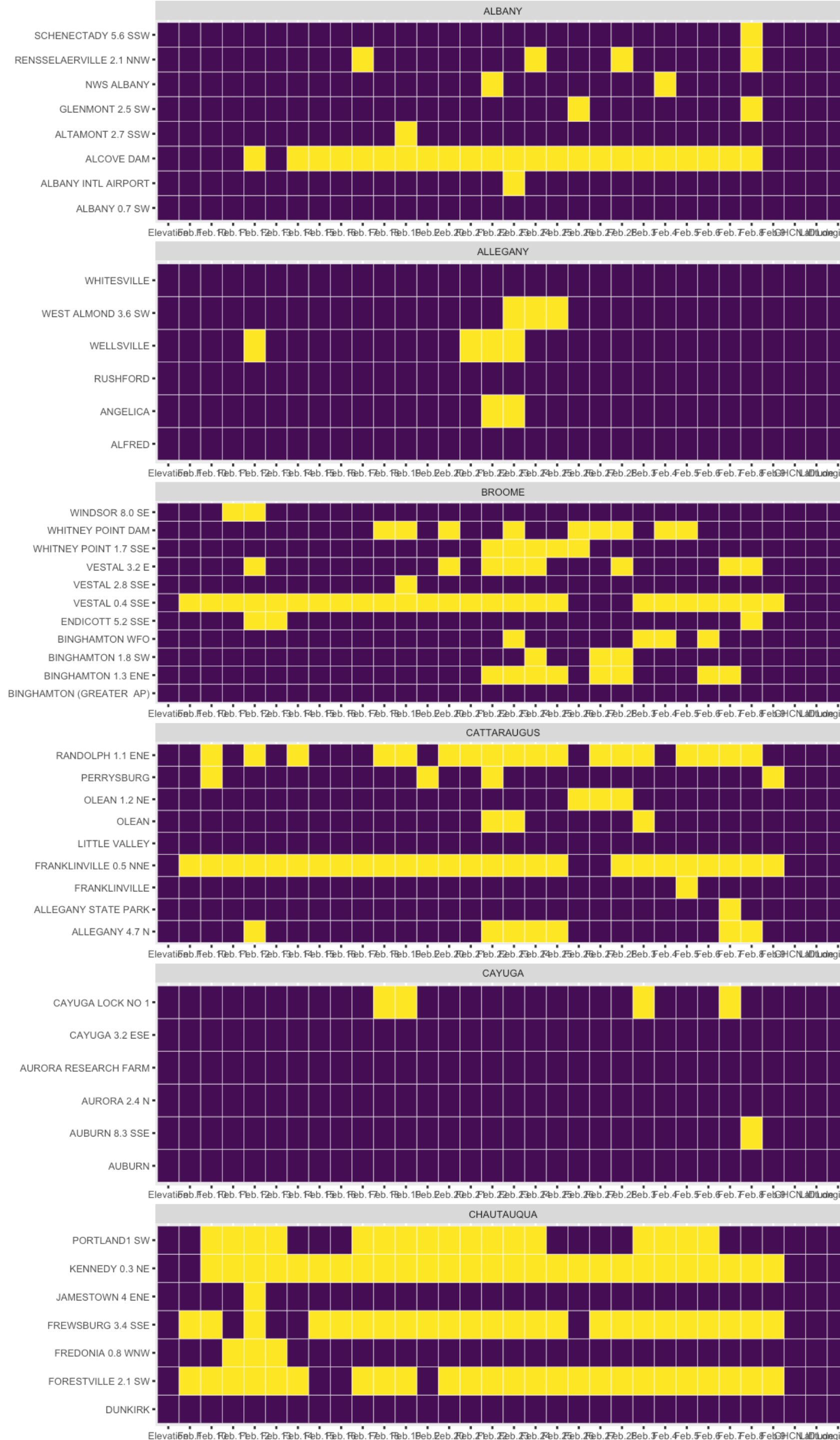
image(missing_data.frame())



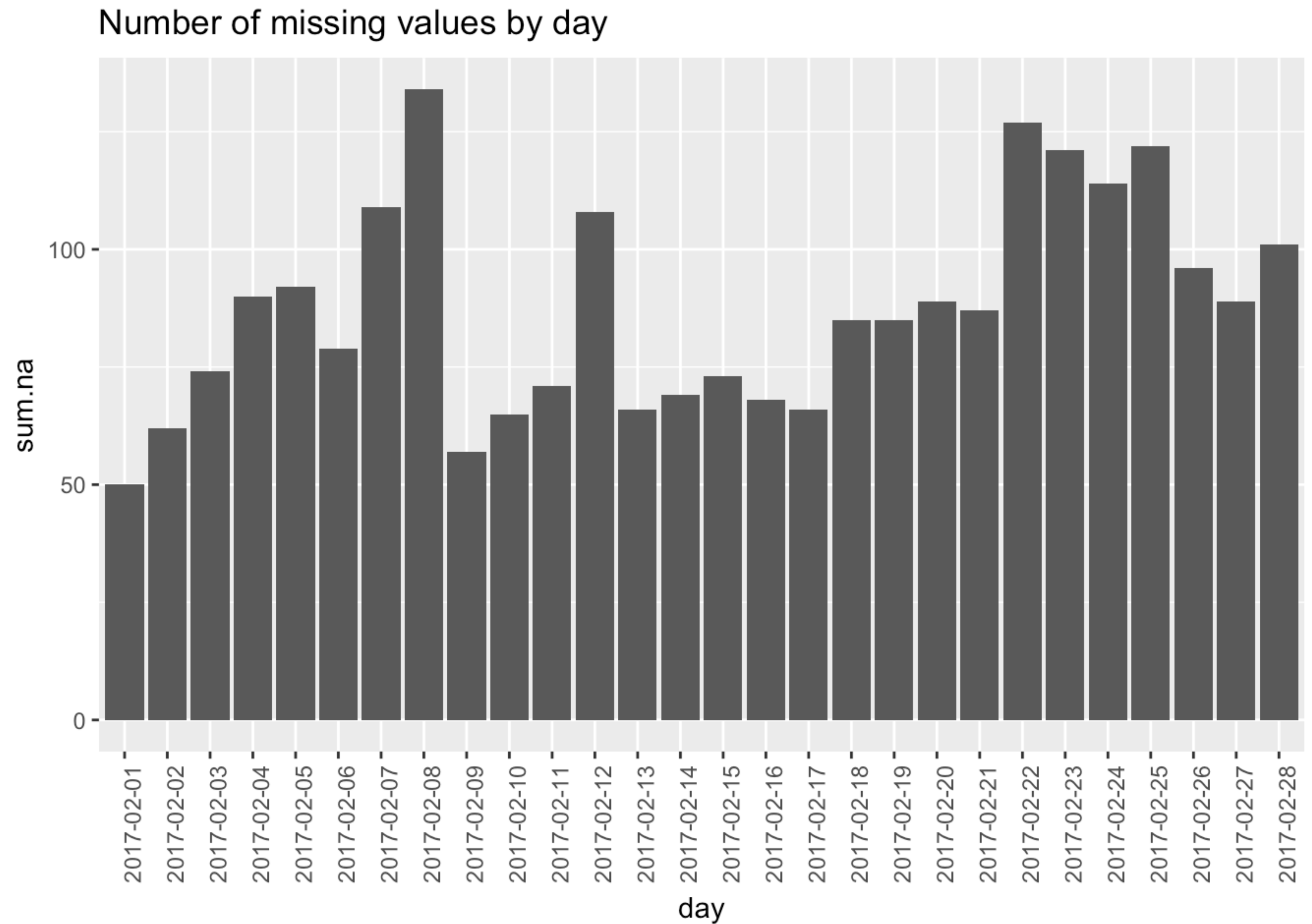
geom_tile()



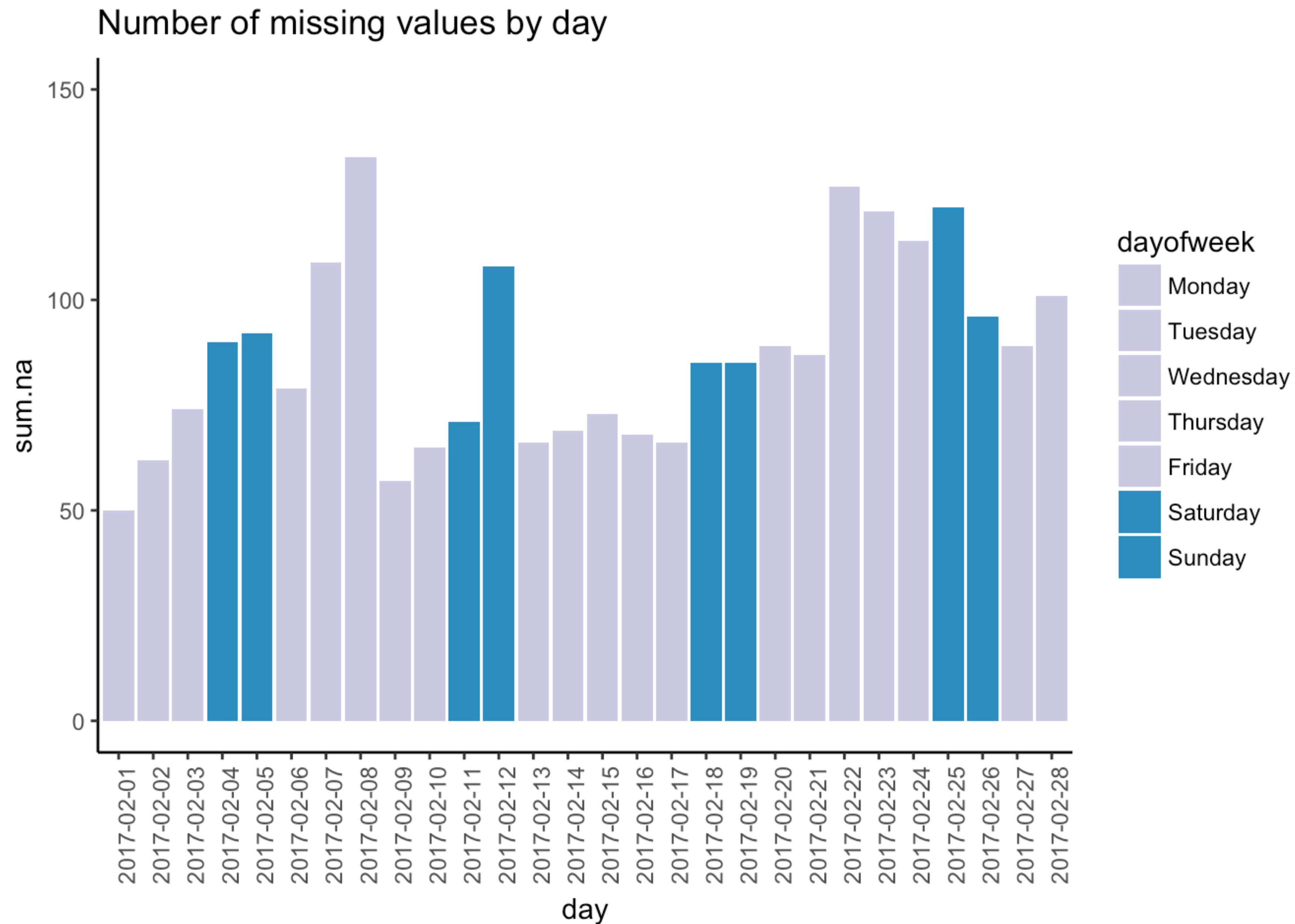
geom_tile()

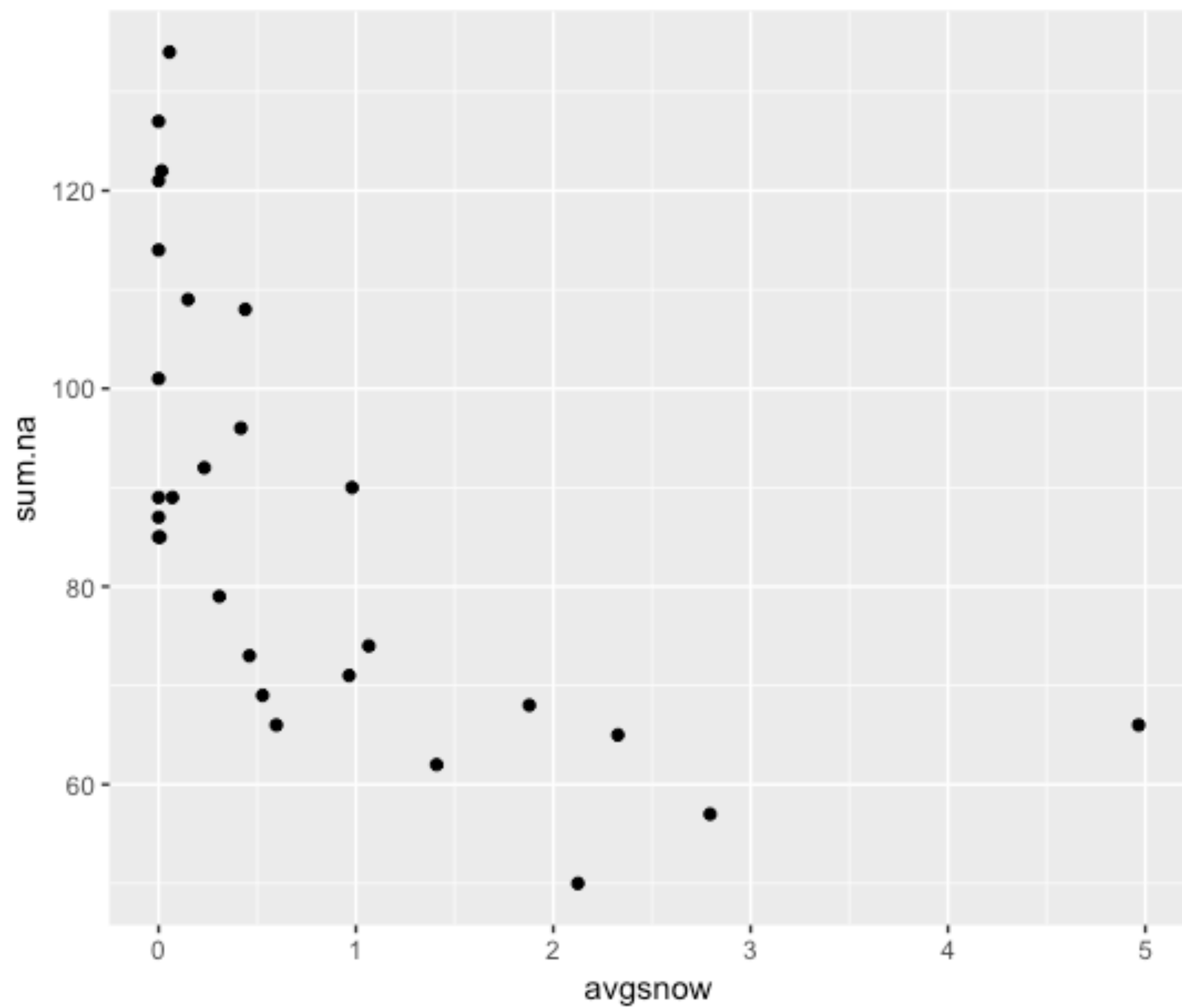


geom_tile()

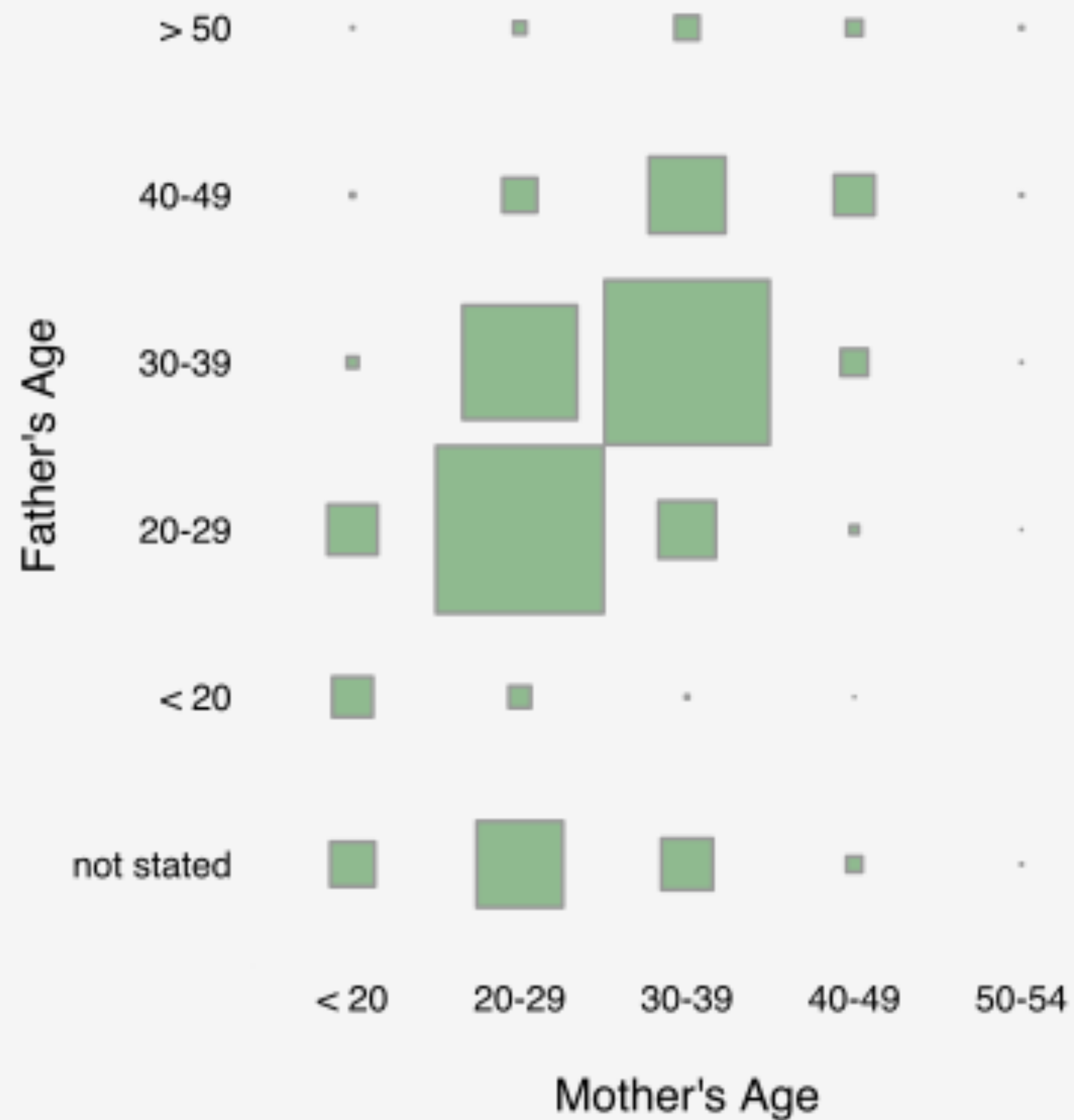


geom_tile()





3,988,076 babies born in the U.S. in 2014:
How old were their parents?



<https://www.ncdc.noaa.gov/snow-and-ice/>

`complete.cases()`

`rowSums(is.na())`

`colSums(is.na())`