Traffic flow on highways in California

- Loop detectors at 22,000 locations,
- Transmit data every 30 seconds
- Collect 2GB a day, and store 4TB

For each of three lanes,
flow (number of cars) and
occupancy (the proportion of time there was a car over the loop)
were recorded in successive five minute intervals.

We have 1740 such five minute intervals.

Lane 1 is the leftmost lane, lane 2 is in the center, and lane 3 is the rightmost.

Statistical problem:

Occupancy can be viewed as a measure of congestion.

- What is the *distribution* of lane occupancy?
- Do you think that the distribution of occupancy is *symmetric* or *skewed*? Why?
- Are there any unusual features of the distribution?
- How does occupancy in different lanes relate to each other?

Let's go to R and see what we can find.

Recap Topics

- Variables – vectors, data frames
- Managing session and variables
- Input/Output and Data – data from web, data to files
- Graphics – devices for display

Variables

Variables have a name and a value.

To access the value we use the name. Variables allow us to

- Store state on the computer
- Store a value without needing to recompute it
- Write a general expression, e.g. sqrt(aˆ 2 + bˆ 2)
- Reduce redundancy (and mistakes)

```
 n = 10
x = rnorm(n)
sum(x) / n
```

Managing Sessions and Variables

We can manage our variables with R functions

- List all variables
  objects()
- Remove one or more variables
  rm(x, y)
- Save variables for future use
  save(x, y, z, file = “myfile.rda”)
- Restore variables
  load(“myfile.rda”)
- Alternatively, an entire workspace may also be saved, and it will be automatically loaded when you start R up again.
  $>$ **q()**
  Save workspace image? [y/n/c]:
  But it keeps EVERYTHING!

Variable Names

Variable Names must follow some rules:

- May not start with a digit or underscore
- May contain numbers, characters (upper and lower case), and some punctuation, period . and underscore _ are okay,
  but most other other are not, e.g. commas, quotation marks, and # are not.
- Case-sensitive, so **x** and **X** are different.
- Use meaningful names.
- Avoid names that have a meaning in R, e.g. function names such as **c**, **t**, **s**, **.C**

Managing Sessions and Variables

- Keeping track of the code you write:
  - To see the code that you have executed in the R session
    history(max.show = Inf)
    savehistory(”myCode.R”)
  - To evaluate code that you have written and saved in a file
    source(”myRevisedCode.R”)
- To get help with functions, begin your session by starting the help browser:
  help.start()
  then when you need specific help on a function you can ask for it as follows
  help(plot)
  or
  ?plot