Data Technologies

R and Databases

Background Reading

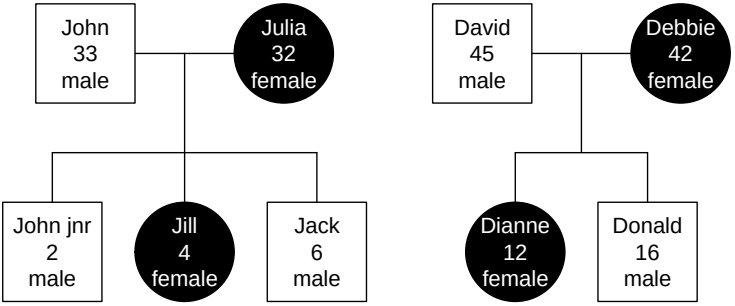
Section 7.9 of “Introduction to Data Technologies”

The ”R Data Import/Export” Manual.

Databases

- tables, rows, and columns
- data types, primary keys, foreign keys

Hierarchical Data



Hierarchical Data

```
John,33,male,Julia,32,female,Jack,6,male
John,33,male,Julia,32,female,Jill,4,female
John,33,male,Julia,32,female,John jnr,2,male
David,45,male,Debbie,42,female,Donald,16,male
David,45,male,Debbie,42,female,Dianne,12,female
```

Hierarchical Data

```
<family>
  <parent gender="male" name="John" age="33" />
  <parent gender="female" name="Julia" age="32" />
  <child gender="male" name="Jack" age="6" />
  <child gender="female" name="Jill" age="4" />
  <child gender="male" name="John jnr" age="2" />
</family>
<family>
  <parent gender="male" name="David" age="45" />
  <parent gender="female" name="Debbie" age="42" />
  <child gender="male" name="Donald" age="16" />
  <child gender="female" name="Dianne" age="12" />
</family>
```

Hierarchical Data

```
person_table (ID [PK], gender, name, age)


1   male    John 33
2 female   Julia 32
3   male    Jack  6
4 female    Jill  4
5   male John Jr  2
6   male   David 45
7 female  Debbie 42
8   male  Donald 16
9 female  Dianne 12
```

Hierarchical Data

```
family_table (parent [PK] [FK person_table.ID],
              child  [PK] [FK person_table.ID])
```

```
1 3
2 3
1 4
2 4
1 5
2 5
6 8
7 8
6 9
7 9
```

Database Design

- Entities and attributes
- Relationships
- Normalisation

R and Large Data

- R holds all data in memory *and* copies objects
- R can communicate with database management software
- Several packages allow you to keep the data on disk instead: filehash, biglm, ff.

Assignment 4

- The `reshape` package
- The `merge()` function

The Project Database

```
              VARIABLE : Mean TS from clear sky composite (kelvin)
              FILENAME : ISCCPMonthly_avg.nc
              FILEPATH : /usr/local/fer_dsets/data/
              SUBSET   : 24 by 24 points (LONGITUDE-LATITUDE)
              TIME     : 16-JAN-1995 00:00
               113.8W 111.2W 108.8W 106.2W 103.8W 101.2W 98.8W  ...
                 27     28     29     30     31     32     33    ...
36.2N / 51:  272.7  270.9  270.9  269.7  273.2  275.6  277.3  ...
33.8N / 50:  279.5  279.5  275.0  275.6  277.3  279.5  281.6  ...
31.2N / 49:  284.7  284.7  281.6  281.6  280.5  282.2  284.7  ...
28.8N / 48:  289.3  286.8  286.8  283.7  284.2  286.8  287.8  ...
26.2N / 47:  292.2  293.2  287.8  287.8  285.8  288.8  291.7  ...
23.8N / 46:  294.1  295.0  296.5  286.8  286.8  285.2  289.8  ...
...
```

The Project Database

```
date_table (date [PK], month, year)


location_table (ID [PK],
                longitude, latitude, elevation)


measure_table (location [PK] [FK location_table.ID],
               date     [PK] [FK date_table.date],
               cloudhigh, cloudmid, cloudlow,
               ozone, pressure, surftemp,
               temperature)
```

The Project Database

```
1 1995-01-16 26.0 34.5  7.5 304.0 835.0 272.7 272.1
2 1995-01-16 23.0 32.0  7.0 306.0 810.0 270.9 270.3
3 1995-01-16 23.0 32.0  7.0 306.0 810.0 270.9 270.3
4 1995-01-16 17.0 29.5  7.0 294.0 775.0 269.7 270.9
5 1995-01-16 19.5 33.0 11.0 308.0 795.0 273.2 271.5
```