12.2.1

1. Table 1, each column is a single variable and each row is a separate observation. Table 2, the cases and population has been gathered together with a “type” key. Table 3, Cases and population are both shown in a single column. Table 4, population and case are in separate tables.

2.

a. table2 %>% spread(type, count) %>% mutate(rate = cases/population)

b. inner\_join(gather(table4a,year,value=cases,`1999`,`2000`), gather(table4b,year,value=population,`1999`,`2000`),by=c('country' = 'country', 'year'='year')) %>% mutate(rate=cases/population)

3. You first need to spread the data. table2 %>% spread(type, count) %>% ggplot(aes(x=year, y=cases)) + geom\_line(aes(group=country),colour = "grey50") + geom\_point(aes(colour = country))

12.3.3

1. Variable types can change

2. You need the tick marks around the columns `1999` and `2000`

3. There are two rows with “age” for Phillip Woods. You need to add another column with the number of the observation

4. You need to gather by the gender

12.4.3

1. Extra decides what to do if there are too many pieces. Fill decides what to do if there aren’t enough pieces.

2. Remove will get rid of the original column. FALSE will keep it.

3. Extract uses a regular expression to find the groupings. Separate has 3 variations because you need to figure out what to separate by. Unite only needs one because there is only one way to combine all the columns into 1.

12.5.1

1. Separate fill will replace missing values with whatever you declare. Complete allows you to have multiple values to replace NAs with.

2. It determines whether you use the value above or below the missing cell to replace the NA.

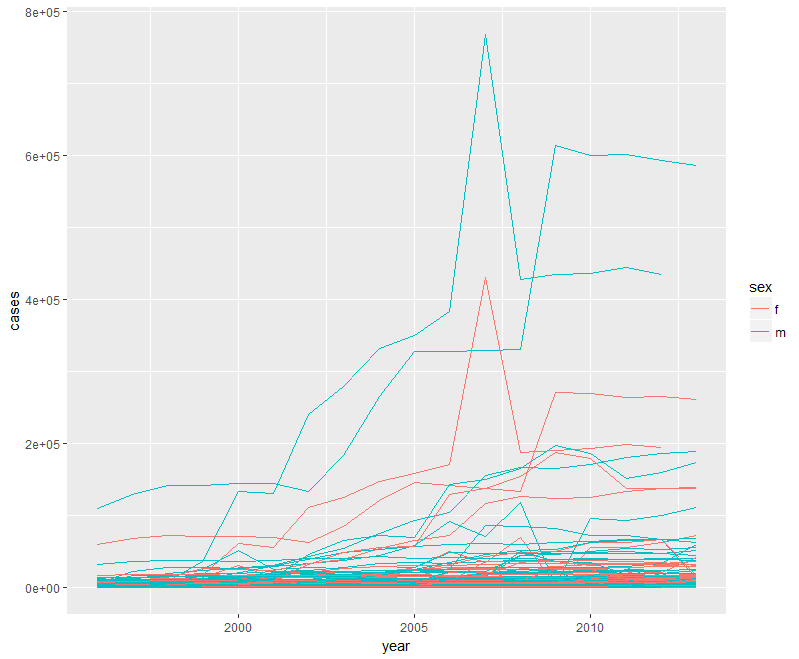
12.6.1

1. Possibly? How did all the zeros get generated in the data?

2. Separate will return a warning and give some wonky results.

3. Group by country to figure it out. Confirmed.

4.



13.2.1

1. Need origin, dest, longitude, and latitude. Merge flights with airports.

2. Origin in weather matched with faa in airports.

3. year, month, day, hour, origin matches with year, month, day, hour, and dest.

4. Create a table of the special dates.

13.3.1

1. mutate(flightID = row\_number())

2.

1. playerID, yearID, stint

2. year, sex, name

3. lat, long, year, month

4. id

5. No primary key.

3.

13.4.6

1. group\_by(flights,dest) %>% summarize(delay=mean(arr\_delay, na.rm=TRUE)) %>% inner\_join(airports, by=c(dest='faa')) %>% ggplot() + borders(aes(x=lon,y=lat,color=delay),'state') + geom\_point() + coord\_quickmap()

2. left\_join(flights, airports, by = c(dest = ‘faa’)) %>% left\_join(airports, by = c(origin = ‘faa’)) %>% head()

3. Nope

4. Minor correlation with precipitation

5. There were a bunch of storms.

13.5.1

1. AA and MQ don’t have tail numbers

2. Create a table that groups by talinum, filter it to where the count is > 100. Then join that back onto flights.

3. semi\_join(fueleconomy::vehicles,fueleconomy::common,by=c('make','model'))

4.

5. Flights that go to an airport that is not in FAA list. The second one are airports that don’t have flights.

6. No relationship like that.