# ĮVADAS

Kad nepasimestume, net ir savo pačių koduose mes laikomės tam tikros tvarkos. Trumpai supažindinu:

Kiekvienam projektui skiriam atskirą folderį, kuriame būtinai turi būti tokie folderiai:

1. Input
2. Output
3. Plots

Rašant kodą, stengiamės,kad jis būtų tvarkingas ir **su komentarais** (taigi bent jau kiekvienas ciklas bei if turi būti trumpai paiškintas). Kiekvienas projektas dažniausiai turi po kelis R script‘us:

1. Load.R (čia nusiskaitom visus duomenis bei juos susitvarkom)
2. Functions.R (čia susidedam funkcijas, kurias naudojam šiame projekte)
3. Main.R (pagrindinis mūsų kodas, kuriame nusi‘source‘inam Load.R, bei Functions.R ir jau modeliuojam bei forecast‘inam)
4. Dar prie kiekvieno projekto nusiskaityk EMI base.RData, kuriame jau bus dalis reiklaingų funkcijų.

Kaip prieš tai minėjau, pabandyk susikuriti projektą, užsikrauti EMI base. RData ir iš karto pasileisti funkciją template(), jin tau turėtų automatiškai šitus dalykus sudėlioti.

Dar vienas labai svarbus dalykas – niekada nekeisk input file‘ų, jie turi išlikti tokie patys, kokius aš tau atsiunčiu. Na aišku, gali pasikeisti viską į csv file‘us, nes dabar pvz. viską tau sudėjau .xlsx, tai tikrai bus geriau, jeigu juos pasiversi csv ir tada importuosi į R. Tada, ten kartais labai gerai būna pasikeisti csv cells formatą, nes R nevisada teisingai viską supranta, tačiau geriau excel‘yje neprijunginėk jokių stulpelių ir nekeisk duomenų.

# Projektas: road injuries

Mes norim truputį pakeisti savo dabartinę miestų gyventojų skaičiaus prognozavimo metodologiją, ją gerokai praplečiant. Dirbsim su 83 šalių detalių administracinių regionų populiacija, bet pradėsim nuo Rusijos ir Kanados. Išbandysim procesą, metodologiją (gali būti, kad reikės ir perdaryti ką nors, nes matysime, kad neveikia), o po to taikysime visoms šalims iš eilės.

### Mūsų pateikiami duomenys (input):

* Nors renkama kur kas daugiau, input naudosime tik tokius duomenis:
  + 134287 Road Injury Accidents
  + 99981 Number of deaths
  + 99971 Number of fatal accidents
  + 99951 For trend
* Papildomai reikia population:
  + National
  + Regional
  + City
* Population share‘ų failas ‚approx‘ atvejams irgi yra paruoštas
* Other inputs
  + Possession of cars (snapshot) – regions/cities/countries
  + Possession of motorcycles – regions/cities/countries
  + Average income
  + Expenditure

### Reikalingi galutinis rezultatas (output):

1. ‚Summary‘ failas pagal anksčiau parašyta scriptą
2. Naudojant originalias / pradines geografines teritorijas:
   1. Tiesiogiai paskaičiuojamas Road injury accidents per 100,000 people
   2. Modelis 1. Road injury įvertis pagal Fatal accidents arba deaths
3. Excel failiukas:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | City | Region | Country | Model Choice |
| CN21 | Original | Original | Original | - |
| CN22 | NA | NA | Original | - |
| MY01 | NA | Original | Original | Possessions |
| MY02 | NA | Based on fatal | Original |  |

1. Sprendimai, kur kokį modelį naudojame užpildyti NA:
   1. Kur tik šalies duomenys – gal paliekam NA
   2. Kur yra bent regionų duomensy – tada based on possession of cars, motorcycles, income, population size (jei population dydis ką nors lemia)
2. Galutinis rezultatas (csv)
3. GALBŪT grafikai kiekvienai šaliai:
   1. Iki normavimo
   2. Po normavimo

### Prognozės

Nedarome