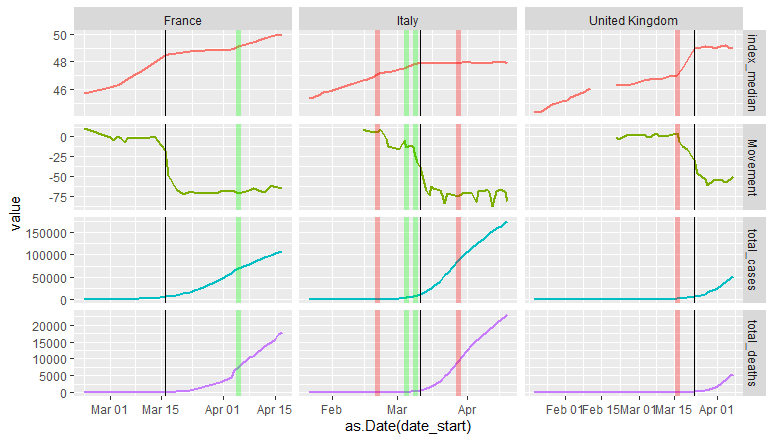
# Things to do in CoronaNet dataset:

* **Identification** of proper lockdowns:
  + ‘**type’** features both ‘**Curfew’** (which it is in the case of France, for example) and ‘**Quarantine/Lockdown’** (Italy, Germany, UK…). The difference between those two seems a bit vague, Curfew seems to be stronger; both will have to be included.
  + However, **mandatory lockdowns also include fairly liberal situations**, like the UK between 16/03 and 23/03, where even pubs were open.
  + Best match with our conception of lockdown will probably come up if we **include closure of non-essential business, closure of schools** and such thingsl; at least in the UK, this leads to 23/03 pretty clear cut.
* **Coding** of lockdown variables:
  + For visualisation and analysis, a **dummy that equals 1 for every day during the lockdown** period when the measure is in place would be useful; again, requires clear definition of what constitutes a lockdown.
* **Interpolate** missing dates:
  + For index, use last reported value
  + For other values, probably right\_join is the better solution (if other dataset has all dates)
* Import most **recent dataset**

Some issues mentioned above visible below:

* Red lines are national mandatory lockdowns as recorded in CoronaNet
* Green lines are national voluntary lockdowns
* Black lines are lockdowns as recorded by Orestis
  + Match is poor so far
* Interpolation needed (ggplot does that automatically except for UK in Feb, as there’s a NaN entry at beginning of gap).



# Analysis:

* Create **Impulse Response Functions** (or multiplier analysis) for intervention dummies: What is the reaction of X (mobility, infections…) to lockdowns (or other measures); trajectory.
* The resulting coefficients can be used on LHS of further analysis: e.g., is reaction stronger in collectivist, authoritarian context etc.
* Based on **VAR or ARDL**?
* VAR:
  + Pro: Can explicitly model endogeneity between variables (case numbers, deaths, mobility, policy index, e.g.)
    - Structure of this interaction may be quite intricate and changing over time – series start off exponential, then turn linear, and lags may be too long for some of these things to be picked up by VAR
  + Con: Quite demanding on data and time series are still fairly short, complicated computation of impulse response for exogenous variables (such as lockdown dummy)
* ARDL:
  + Easier to handly, may suffice if we’re only interested in simple interactions (not trying to model whole case numbers, death, lockdown nexus)
  + Impulse response easier to compute
  + Not so demanding on data (less parameters, eats fewer degrees of freedom)