**!! Project Pipeline !!**

（有些file不需要手动跑，因为有source，加粗的都是需要手动跑的）

Utils 有一些models需要用的functions

->models

-> setting 初步初始化各个variables和function

**Experiment\_models.r** 生成不同parameter的forecast

-> **experiments.rmd** 比较不同parameter的forecast并找到最佳parameter

-> **setting** 把experiment得到的parameter写到defaults里面

-> **run\_prediction** 用最佳parameter生成forecast

**Utils**

Census: 导入数据

Delays: EpiEstim和EpiNow2的functions

Diagnostics：用于testing.rmd。train and predict utils

Eval：所有evaluation相关的r code（evaluation.rmd, experiments.rmd）都用得到的functions

Predict：用于testing.rmd。Train and predict

Prior: EpiEstim和EpiNow2的prior

**Helper**

Plotting：selected\_covid\_data\_us中用来plot data

Prep：用于testing.rmd。一些零碎的helper functions

**Setting**

基础的variable和function的declare。

Defaults：除了基础的declare，把models folder中定义的五个run\_prediction\_us functions整合到train\_and\_predict中，并定义run\_prediction.r中各个model所使用的parameter。

Naming：

**Data**

Data ploting and preprocessing

**Models**

**Gaussian process**

1. Why use log-transformed

**Prophet**

**SARIMA**

**Epiestim**

**Epinow2**

**Experiments**

准备给experiments.rmd用的data

每个model的experiment file都使用多种parameter combinations生成了多个forecasts。这些不同parameters的forecasts将在experiments.rmd中被计算CRPS、进行比较、并挑选出各个model的最佳parameter。

1. 为什么SARIMA用（1,0,1）（1,1,1）（2,0,2）（2,1,2）（2,2,2）
2. What is rds data, why use this type

Experiments.rmd

用于hyper-parameter tuning。比较各个models不同parameters的CRPS

（只evaluate CRPS，并没有计算其他secondary performance metric，因为这一步只是用来挑各个model的parameter。而evaluation.rmd应该是用来彻底、详细地evaluate各个模型（最佳parameter）的表现。）

loading pre-computed forecasts, scoring them (with CRPS), summarizing the results, and then visualizing how different model variants (e.g. different parameter settings) perform across 1- and 2-week–ahead horizons.

**Run prediction.r**

使用setting/defaults.r中的parameter，给每个models和每个stats生成forecasts并保存。

**Evaluation**

Evaluation

**Testing**

Test functions, debug

Debug内容

1. Delay：

因为incubation period和generation time的结构变了，所以plot\_delay\_distr需要大改。原本可以delay$mean，现在得delay$parameters$meanlog$parameters$mean等等

1. Testing：

Espistim的config-si的FUN没有define

1. Epinow2

Epinow()的Input parameters整个大变样了，需要把整个project中所有相关的都改掉

Generation\_time, rt, pt, data

For gt, ls\_min has been deprecated, is it necessary?

Epinow2()的delay\_opts()也变了，原本是能输入格式相同的reporting\_delay和incubation\_period，然后自动相加。但是更新之后的incubation\_period格式和reporting\_delay不一样了，delay\_opts()也只支持输入一个distribution了。所以调整了make\_reporting\_delay()，让得到的reporting\_delay和incubation\_period格式相同。

1. Epistim

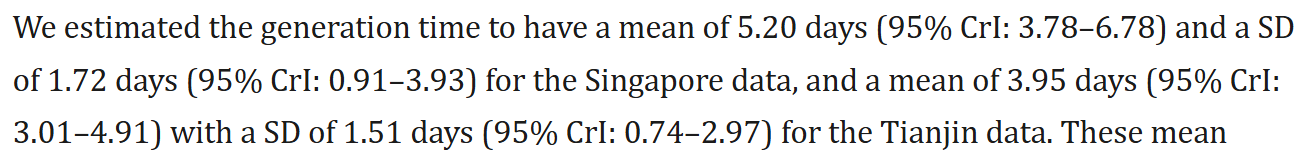
Calculate mean and sd based on shape and rate

Change the parameters in make\_config

原本：

Incubation\_period: Lauer et al. (2020)

Generation\_time: Ganyani et al. (2020) in mean and sd



Generation\_time on paper: Hart et al. (2021)

Mean: mean=4.2 (95%-CrI: 3.3-5.3)

SD: mean=4.9 (95%-CrI: 3.0-8.3)

现在:  
Incubation\_period: Lauer et al. (2020)

Generation\_time: Ganyani et al. (2020) in shape and rate

告诉prof：

1. make\_config应该不止需要mean\_si和sd\_si
2. 我不需要estimate generation time的mean and sd了，因为我在ganyani的原paper上找到了mean and sd的95% CI。但我不知道paper用的是天津的还是新加坡的数据
3. Additional: paper的citation有点问题。里面说用的是Hart et al. (2021)的generation time的数据，实际上引用的是Ganyani et al. (2020)。而代码中调用的也是Ganyani et al. (2020)
4. Epinow2
5. Incubation period + reporting delay
6. ~~Gt\_opts~~: 用的就是example generation time，但也有可能就是因为从mean, sd -> shape, rate导致的问题
7. ls\_min
8. ~~rt\_opts~~: 用的就是原来的数据，只是封装在了一个list里面
9. Delay\_opts到底是什么情况