ИНСТРУКЦИЯ

Пошаговый запуск байесовской модели
прогнозирования кори (EpiNow2)
для Кыргызской Республики

Файл скрипта: epinow2_measles_ru.R

Язык вывода: русский

Содержание

Часть I. Подготовка 3
1. Установка программ 3
2. Структура папок проекта 3
3. Подготовка входных файлов 3
4. Создание R-проекта в RStudio 4
Часть II. Запуск скрипта (шаг за шагом) 6
5. Общий принцип пошагового запуска 6
6. Шаг 1 — Подключение библиотек 7
7. Шаг 2 — Настройки источника данных 8
8. Шаг 3 — Эпидемиологические параметры 8
9. Шаг 4 — Загрузка данных из Excel 9
10. Шаг 5 — Расчёт задержек регистрации 10
11. Шаг 6 — Настройка модели 12
12. Шаг 7 — Национальная модель 12
13. Шаг 8 — Региональная модель 14
14. Шаг 9 — Загрузка вспомогательных функций 14
15. Шаг 10 — Экспорт национальных результатов 16
16. Шаг 11 — Экспорт региональных результатов 17
Часть III. Частые ошибки и итоговый чек-лист 19
17. Частые ошибки и что делать 19
18. Итоговый чек-лист: что должно получиться 19

ЧАСТЬ I. ПОДГОТОВКА

1. Установка программ

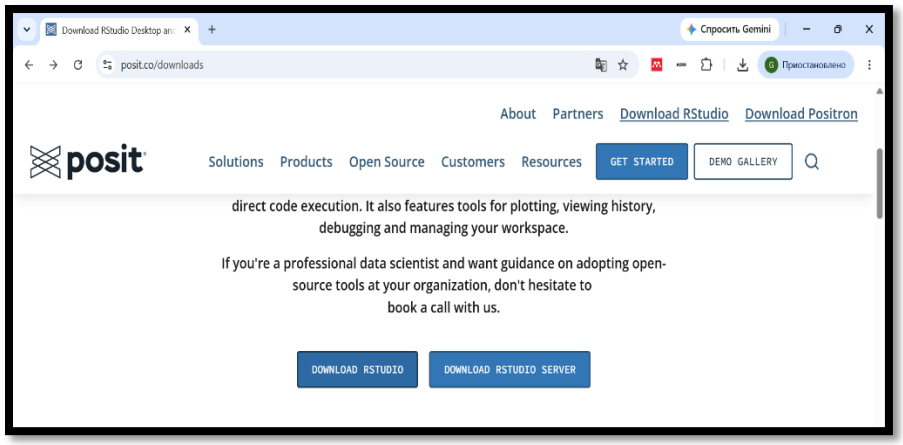
На компьютере должны быть установлены две программы (обе бесплатные):

> ⚠ Порядок установки: сначала R, затем RStudio. RStudio сам найдёт установленный R.

Программа	Где скачать	Зачем нужна
R (4.3+)	https://cran.r-project.org	Язык, в котором работает модель EpiNow2



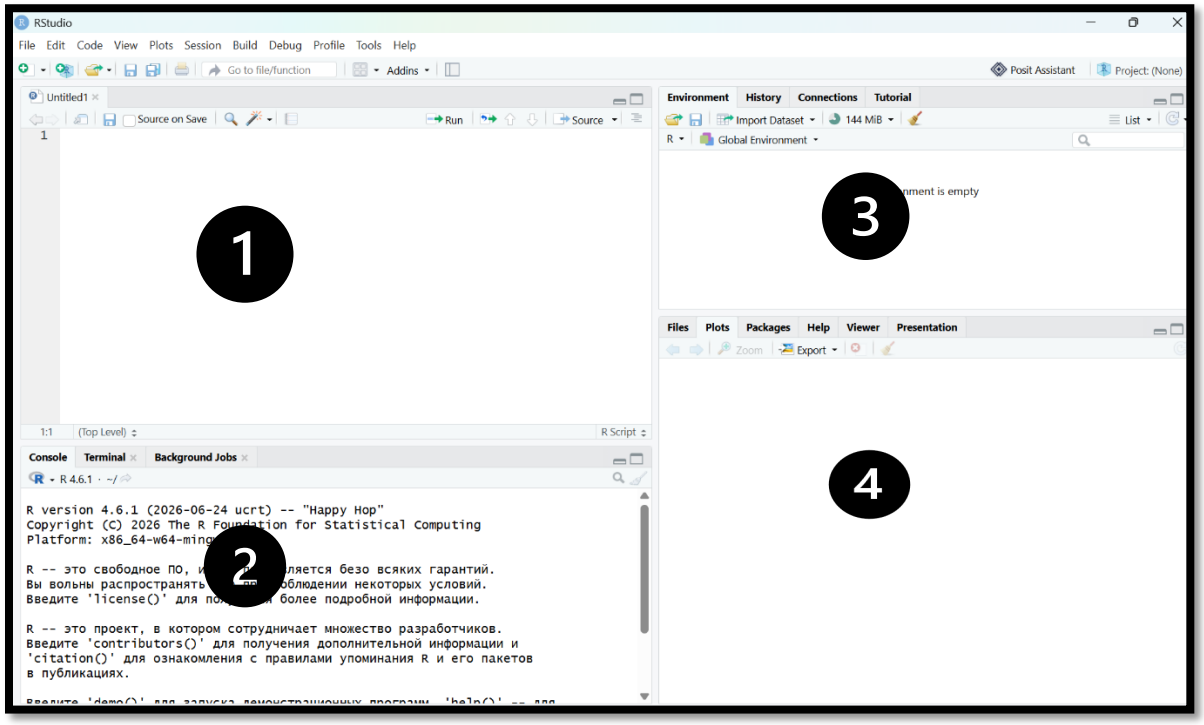
Программа	Где скачать	Зачем нужна
RStudio Desktop	https://posit.co/downloads/	Удобный редактор для работы с кодом R



Запускайте только RStudio. Откройте меню «Пуск» (в левом нижнем углу экрана) или нажмите клавишу Win на клавиатуре. Начните вводить слово RStudio. Нажмите на синий квадратный ярлык с белой буквой R.

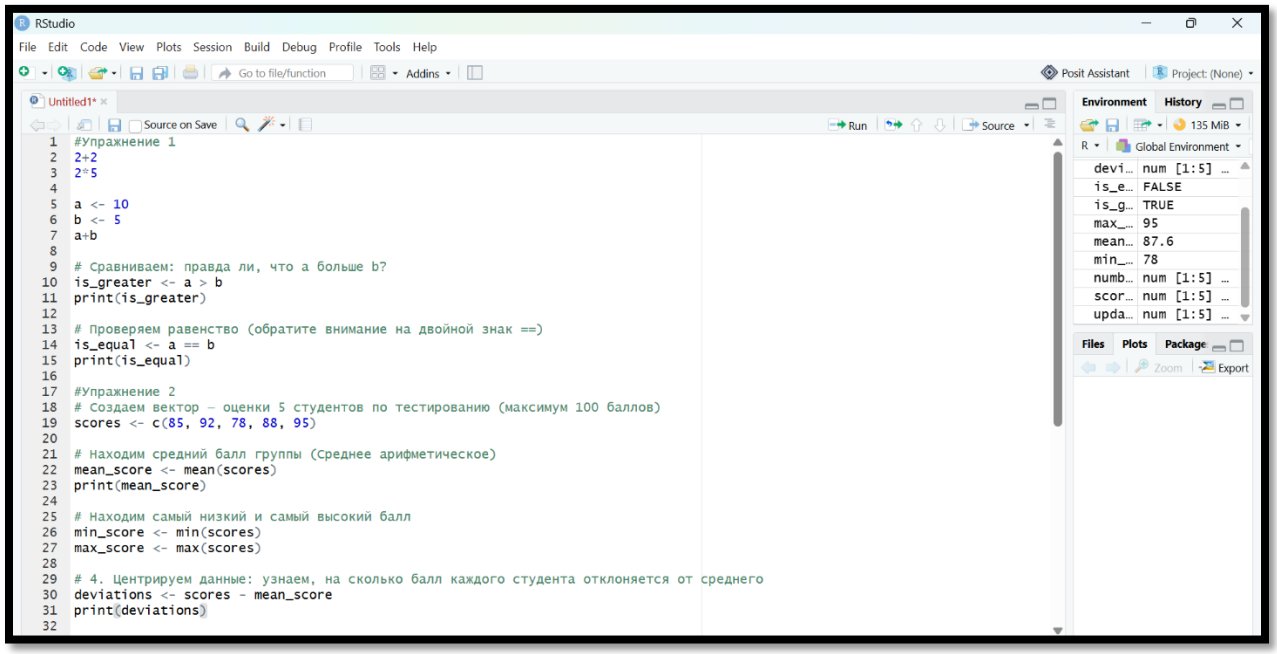
Важно: Саму программу R (базовый R) запускать с рабочего стола не нужно. RStudio автоматически найдет R и станет для него красивым и удобным «рабочим кабинетом».

Интерфейс RStudio.



1. **Source (Редактор скриптов)**: левый верхний угол. Здесь вы пишете, редактируете и сохраняете свой код.
2. **Console (Консоль)**: левый нижний угол. Сюда выводятся результаты вычислений. Здесь также можно быстро выполнять команды без сохранения.
3. **Environment / History (Окружение)**: правый верхний угол. Показывает загруженные данные, созданные переменные и историю выполненных команд.
4. **Files / Plots / Packages / Help**: правый нижний угол. Многофункциональная панель для просмотра файлов, графиков, списка библиотек и справок по функциям.

Выполним несколько команд для ознакомления с интерфейсом RStudio.



2. Структура папок проекта

Минимальные аппаратные требования: 8 ГБ оперативной памяти (лучше 16 ГБ), процессор с 4 ядрами, 2 ГБ свободного места на диске. Интернет нужен при первом запуске для скачивания пакетов.

Создайте папку проекта и внутри неё две подпапки:

```
measles_forecast/            ← корневая папка проекта
├── data/                      ← сюда кладём Excel-файл с данными
├── inst/                      ← сюда кладём ref.csv (справочник задержек)
└── epinow2_measles_ru.R       ← скрипт (этот файл)
```

ℹ️ Папки int/ (кеш моделей) и output/ (результаты) скрипт создаст автоматически. Создавать их вручную не нужно.

3. Подготовка входных файлов

3.1. Файл данных (.xlsx)

Файл Excel с данными эпиднадзора по кори. Может иметь **любое имя**, но должен быть один в папке `data/`. Требования:

Параметр	Требование
Формат	.xlsx (Excel 2007 и новее)
Имя листа	`Новое приложение 4` (точно так, с пробелами)
Столбец 3	Название региона (текст, кириллица)
Столбец 5	Дата регистрации случая (формат даты)
Столбец 6	Статус: `лаб.подт.`, `клин.корь` или `эпид.связь`

⚠ Если в папке data/ окажется несколько .xlsx-файлов, скрипт, скорее всего, завершится с ошибкой ещё до чтения данных — readxl не умеет сам выбирать между несколькими путями. Убедитесь, что в data/ лежит ровно один нужный файл, прежде чем запускать Шаг 4.

3.2. Справочник задержек регистрации (ref.csv)

CSV-файл с информацией о задержке от заражения до регистрации по каждому региону. Имя файла: строго `ref.csv`. Кладётся в папку `inst/`.

Столбец	Что содержит
`region`	Название региона (точно как в xlsx)
`pop`	Население региона
`rep_avg`	Средняя задержка регистрации (дни)
`rep_plus_minus`	Разброс ± дней (95% полуинтервал)
`max`	Макс. задержка (дни). Необязательное поле.

Пример:

```
region,pop,rep_avg,rep_plus_minus,max
г.Бишкек,1100000,5,3,30
Чуйская,980000,7,4,30
Ошская область,1400000,10,5,45
```

⚠ Названия регионов в ref.csv должны побуквенно совпадать с xlsx: г.Бишкек (без пробела после точки), Ошская область (не Ош. обл.) и т.д.

4. Создание R-проекта в RStudio

R-проект привязывает RStudio к вашей папке, чтобы скрипт мог найти файлы данных. Создаётся один раз:

1. Откройте RStudio.
2. Меню: **File** → **New Project** → **Existing Directory**.
3. Нажмите **Browse**, выберите папку `measles_forecast`.
4. **Create Project**.

В папке появится файл `measles_forecast.Rproj`. В дальнейшем открывайте проект двойным щелчком по этому файлу.

ℹ️ Почему это важно: без R-проекта RStudio не знает, где искать папки data/ и inst/. Скрипт выдаст ошибку «Не найден .xlsx».

ЧАСТЬ II. ЗАПУСК СКРИПТА (ШАГ ЗА ШАГОМ)

5. Общий принцип пошагового запуска

Скрипт разделён на 11 пронумерованных шагов (Шаг 1 — Шаг 11). Каждый шаг выполняется отдельно, по очереди. Это позволяет:

- видеть результат каждого этапа в консоли;
- обнаруживать ошибки раньше, чем они накопятся;
- при возникновении проблемы перезапустить только один шаг, а не весь скрипт.

Как выполнить один шаг

1. В RStudio откройте файл `epinow2_measles_ru.R` (вкладка Files → щёлкните по файлу).
2. Найдите нужный блок по заголовку (например, `# ШАГ 4. Загрузка и подготовка данных`).
3. Выделите мышью все строки этого блока — от заголовка до следующего заголовка «ШАГ».
4. Нажмите **Ctrl+Enter** (Windows) или **Cmd+Enter** (Mac).
5. Дождитесь: в консоли (нижняя панель) появится результат или сообщение.

ℹ️ Шаги нужно выполнять строго по порядку: Шаг 1, затем Шаг 2 и т.д. Каждый шаг использует результат предыдущего.

⚠️ Нельзя запустить Шаг 10, не выполнив Шаг 7 (национальную модель), — не будет файла с результатами.

6. Шаг 1 — Подключение библиотек

Что делает этот код

Загружает в память все программные пакеты (библиотеки), которые нужны для работы скрипта. Пакет `pacman` автоматически проверяет: если пакет ещё не установлен — скачивает и устанавливает его из интернета. Если уже установлен — просто подключает.

Какие пакеты подключаются

Пакет	Для чего нужен
`here`	Определяет путь к папке проекта автоматически
`readxl`	Читает файлы Excel (.xlsx)
`readr`	Читает CSV-файлы и сохраняет RDS-файлы
`dplyr, tidyr`	Обработка и трансформация таблиц данных
`lubridate`	Работа с датами (преобразование, сравнение)
`ggplot2`	Построение графиков
`patchwork`	Объединение нескольких графиков в одну картинку
`EpiNow2`	Байесовская модель прогнозирования эпидемий

Что вы увидите в консоли

Сообщения о подключении пакетов. При первом запуске — сообщения о скачивании и установке (это может занять 5–15 минут). Если появится вопрос — введите y и нажмите Enter.

⚠ Этот шаг требует интернет-соединения при первом запуске. В дальнейшем интернет не нужен.

7. Шаг 2 — Настройки источника данных

Что делает этот код

Задаёт параметры, по которым скрипт будет искать данные в Excel-файле. Эти значения сохраняются в памяти как именованные настройки (константы). Сам файл на этом этапе ещё не открывается.

Что означает каждая строка

Настройка	Что задаёт
`DATA_SHEET`	Имя листа в Excel, откуда читать данные
`COL_REGION <- 3`	Номер столбца с названием региона
`COL_REGDATE <- 5`	Номер столбца с датой регистрации
`COL_CONFIRM <- 6`	Номер столбца со статусом подтверждения
`CONFIRM_OK`	Какие значения считать подтверждёнными случаями
`WINDOW_DAYS <- 150`	Окно анализа: последние 150 дней от самой поздней даты
`REF_CSV`	Путь к справочнику задержек (inst/ref.csv)

Что вы увидите в консоли

Ничего — настройки просто сохранились в памяти. Ошибок быть не должно.

ℹ Если структура вашего Excel-файла отличается (например, регион в столбце 4, а не 3), измените соответствующее число в коде перед запуском.

8. Шаг 3 — Эпидемиологические параметры кори

Что делает этот код

Задаёт два ключевых эпидемиологических параметра кори, которые модель EpiNow2 использует для расчётов:

Параметр	Что это и зачем
`gen` — генерационный интервал	Среднее время от заражения одного человека до заражения следующего. Для кори: 11.7 дней (SD 3 дня). Модель использует этот параметр, чтобы связать число заражений сегодня с числом заражений в прошлом.
`inc` — инкубационный период	Время от заражения до появления симптомов. Для кори: медиана 12 дней. Модель использует этот параметр, чтобы пересчитать случаи «по дате регистрации» в случаи «по дате заражения».

Что вы увидите в консоли

Ничего — параметры сохранились в памяти. Эти значения взяты из научной литературы и обычно не требуют изменений.

9. Шаг 4 — Загрузка данных из Excel

Что делает этот код

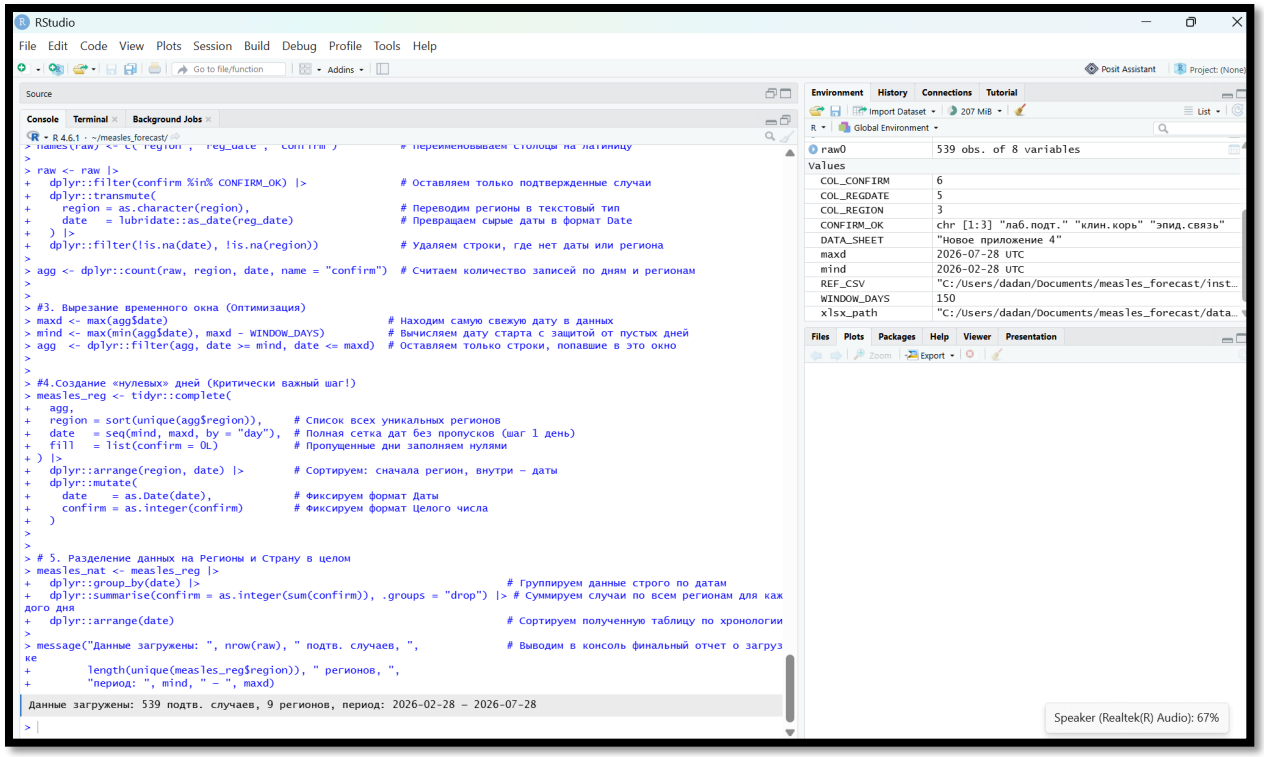
Это самый объёмный блок подготовки данных. Он выполняет следующие действия по порядку:

1. **Находит Excel-файл** в папке data/. Если файла нет — выдаёт ошибку и останавливается.
2. **Открывает лист «**`Новое приложение 4`**» и читает три столбца**: регион (столбец 3), дата (столбец 5), статус (столбец 6).
3. **Фильтрует данные**: оставляет только строки, где статус = лаб.подт., клин.корь или эпид.связь. Остальные строки отбрасываются.
4. **Считает случаи**: подсчитывает, сколько подтверждённых случаев приходится на каждый день в каждом регионе.
5. **Ограничивает период анализа**: оставляет только последние 150 дней (параметр WINDOW_DAYS из Шага 2).
6. **Заполняет пропуски нулями**: если в какой-то день в каком-то регионе не было случаев, ставит 0 (а не пропуск). Это требование модели EpiNow2.
7. **Создаёт национальный ряд**: суммирует все регионы по каждой дате, получая общее число случаев по стране.

Что вы увидите в консоли

Сообщение вида: `Данные загружены: 539 подтв. случаев, 9 регионов, период: 2026-02-28 — 2026-07-28`

Это подтверждает, что данные прочитаны корректно. Проверьте, соответствуют ли число случаев, количество регионов и период вашим ожиданиям.



⚠ Если вы видите 0 случаев или неправильное число регионов — проверьте, что столбцы в Excel соответствуют номерам в Шаге 2, и что названия статусов совпадают с CONFIRM_OK. Частая причина: число регионов в сообщении Шага 4 не совпадает с восемью регионами в словаре переводов (Шаг 9) — почти всегда это опечатка или лишний пробел в названии региона в Excel (например, «Ошская область» и «Ош.область» скрипт считает разными регионами). Сверьте список уникальных названий регионов в Excel с восемью названиями в справочнике ref.csv и со словарём tr$region_names в коде.

10. Шаг 5 — Расчёт задержек регистрации

Что делает этот код

Читает справочник `inst/ref.csv` и рассчитывает, насколько в среднем случай «запаздывает» от момента заражения до регистрации в системе надзора. Это важно: если задержка большая, то сегодняшние данные отражают заражения двухнедельной давности, а не вчерашние.

Рассчитываются два вида задержек:

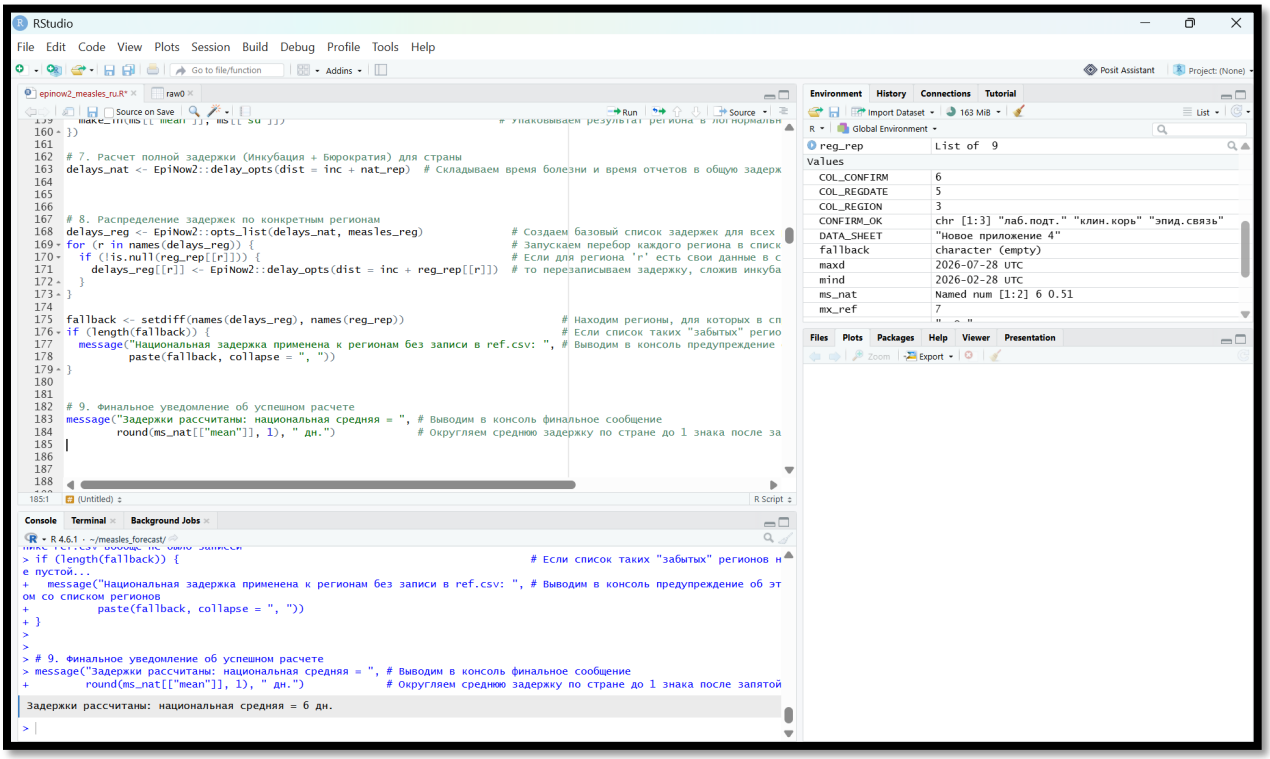
- **Национальная** — усреднённая по всем регионам, взвешенная по населению. Используется для общего анализа по стране.
- **Региональные** — отдельная задержка для каждого региона. Используется в региональной модели. Если для какого-либо региона нет строки в ref.csv, ему автоматически присваивается национальная задержка.

Далее задержка **складывается** с инкубационным периодом (из Шага 3), формируя полную задержку: от заражения до появления в системе надзора.

Что вы увидите в консоли

Сообщение: `Задержки рассчитаны: национальная средняя = 6 дн.` (число будет зависеть от ваших данных в ref.csv).

Если какие-либо регионы не найдены в ref.csv, появится предупреждение: `Национальная задержка применена к регионам без записи в ref.csv: ...`



11. Шаг 6 — Настройка модели

Что делает этот код

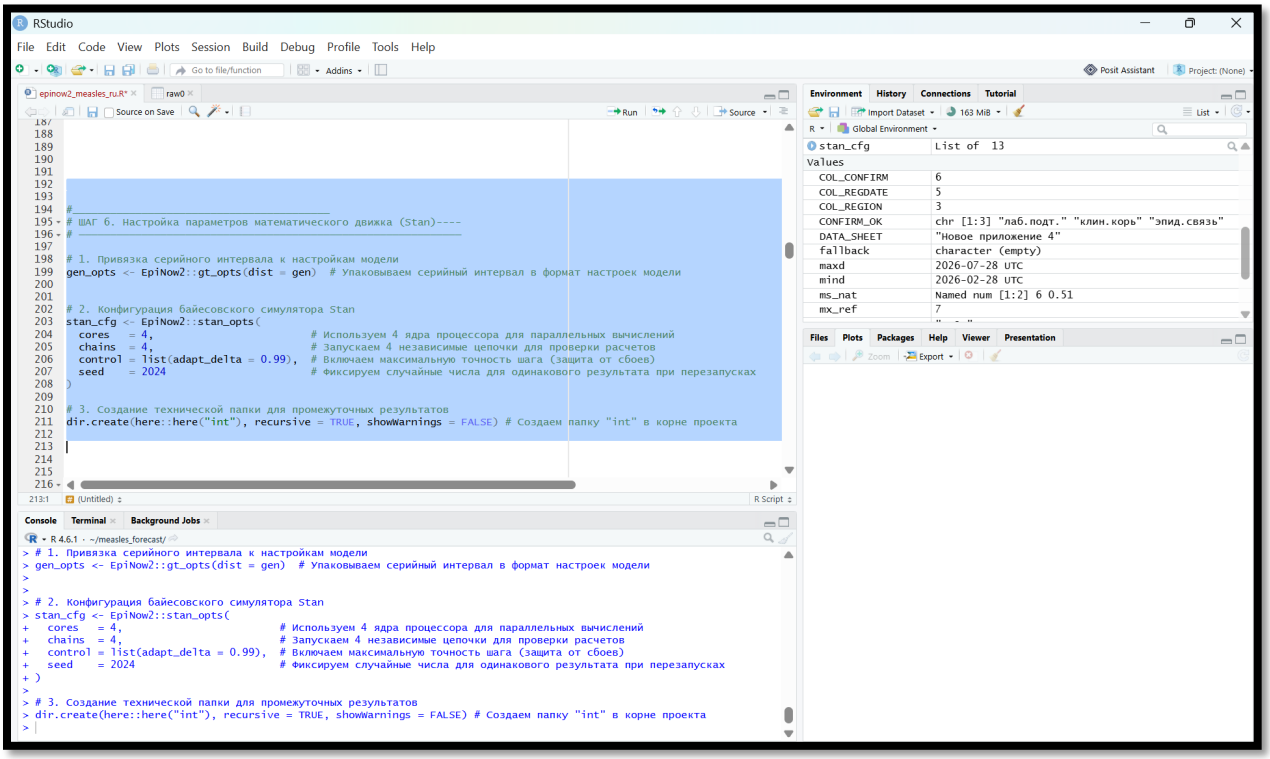
Задаёт параметры байесовского сэмплера Stan — внутреннего вычислительного движка модели EpiNow2:

- `cores = 4` — использовать 4 ядра процессора параллельно (ускоряет расчёт).
- `chains = 4` — запустить 4 независимые цепи Маркова (стандарт для байесовского анализа).
- `adapt_delta = 0.99` — точность сэмплирования (0.99 = высокая, для сложных моделей).
- `seed = 2024` — зерно генератора случайных чисел для воспроизводимости результатов.

Также создаётся папка `int/` для хранения промежуточных результатов (кеша моделей).

Что вы увидите в консоли

Ничего, кроме сообщения о создании папки.



12. Шаг 7 — Национальная модель

ВНИМАНИЕ: этот шаг занимает 30–60 минут!

Что делает этот код

Запускает основную байесовскую модель EpiNow2 для всей страны. Модель:

- берёт ежедневный ряд случаев (из Шага 4);
- учитывает генерационный интервал и задержку регистрации;
- оценивает эффективное репродуктивное число (Rt) на каждый день;
- строит 14-дневный прогноз числа случаев.

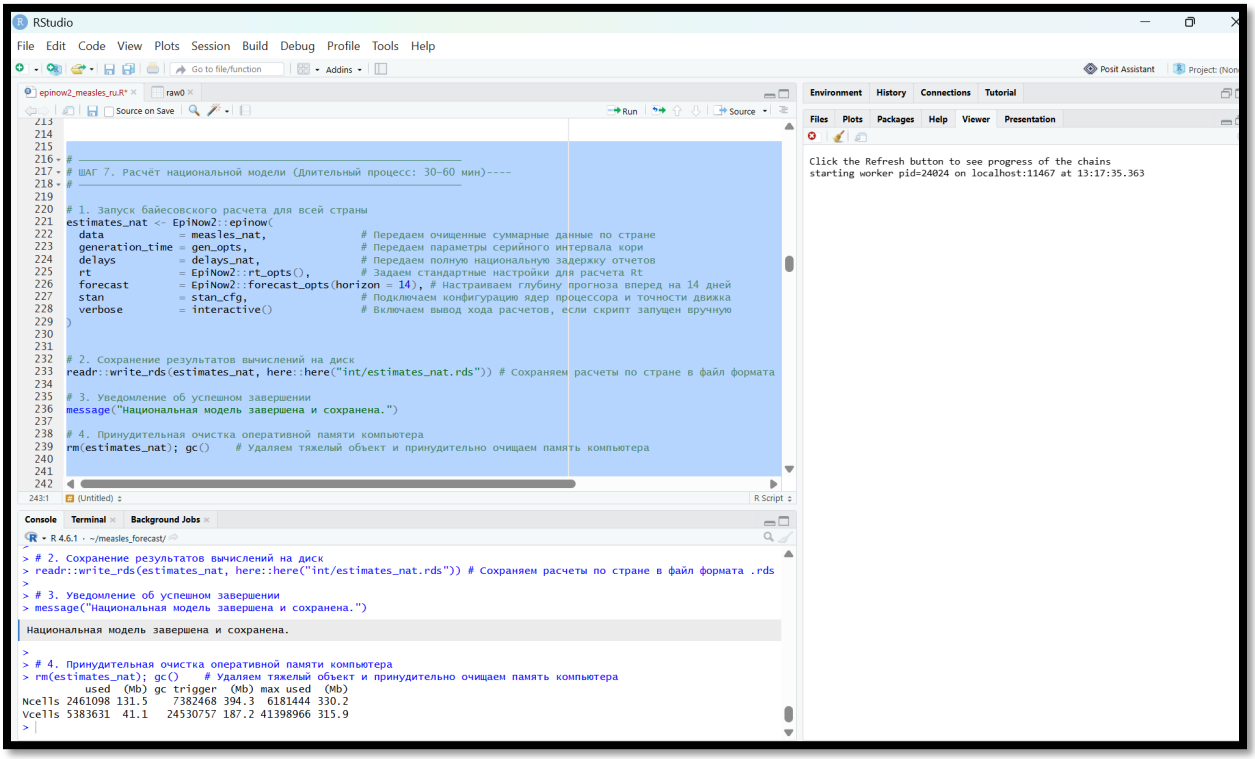
Результат сохраняется в файл `int/estimates_nat.rds` (кеш). После сохранения объект удаляется из памяти, чтобы освободить ресурсы.

Что вы увидите в консоли

Прогресс сэмплирования: номер цепи, процент завершения, скорость (iterations/sec). Это нормальный вывод Stan. Могут быть предупреждения о «divergent transitions» — при небольшом их количестве (<50) это допустимо.

По окончании: `Национальная модель завершена и сохранена.`

⚠ Не закрывайте RStudio и не выключайте компьютер, пока идёт расчёт. Если прервать расчёт, его придётся начинать заново.



13. Шаг 8 — Региональная модель

ВНИМАНИЕ: этот шаг занимает 30–90 минут!

Что делает этот код

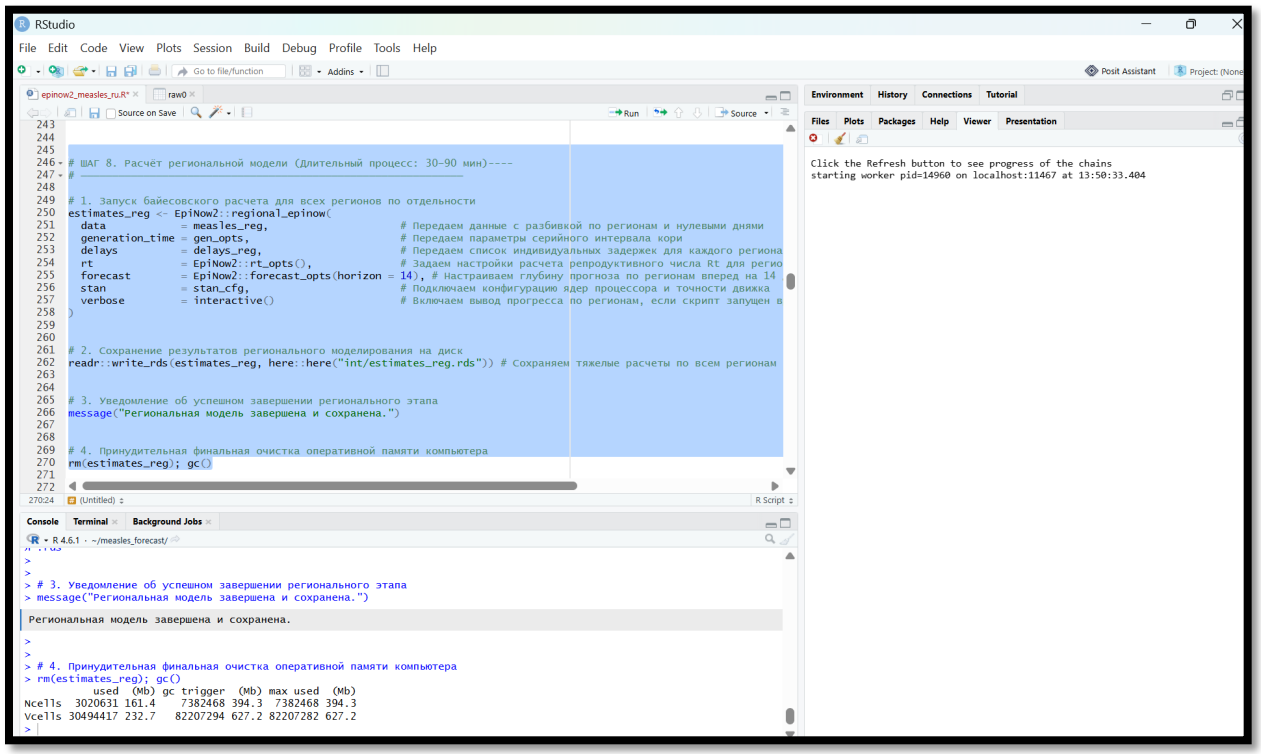
Запускает ту же модель EpiNow2, но отдельно для каждого из 8 регионов. Каждый регион получает свою задержку регистрации (из Шага 5). Модель оценивает Rt и прогноз для каждого региона независимо.

Результат сохраняется в файл `int/estimates_reg.rds`.

Что вы увидите в консоли

Последовательный вывод прогресса для каждого региона: название региона, затем сэмплирование 4 цепей. Это самый длительный этап.

По окончании: `Региональная модель завершена и сохранена.`



14. Шаг 9 — Загрузка вспомогательных функций

Что делает этот код

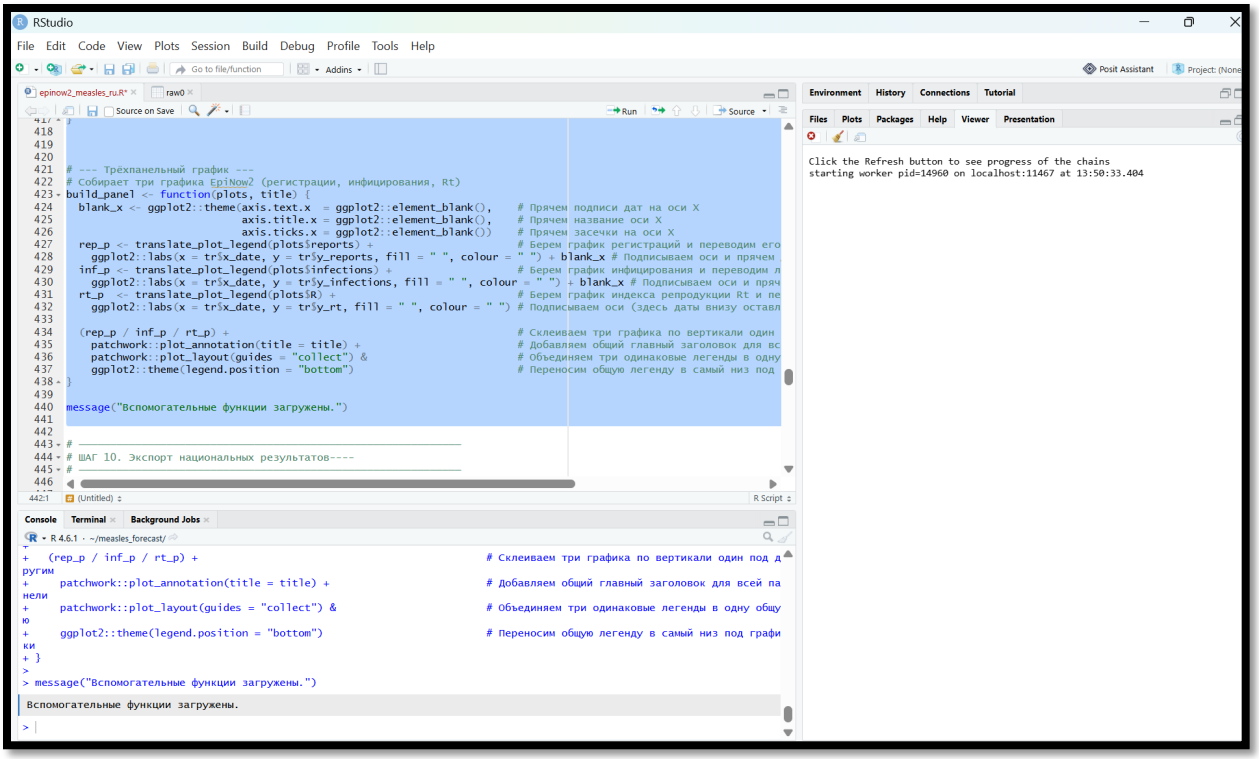
Загружает в память набор функций, которые нужны для Шагов 10–11. К ним относятся:

- **tr** — словарь русских подписей для графиков и таблиц (названия осей, легенд, мер).
- `sanitize_path` — превращает кириллические названия регионов в латиницу для названий папок (Windows не всегда корректно работает с кириллицей в путях).
- `translate_direction` — переводит направления тренда (Increasing → Увеличение и т.д.).
- `translate_measure` — переводит названия мер (Effective reproduction no. → Эффективное репродуктивное число).
- `translate_plot_legend` — переводит легенду на графиках (Estimate → Оценочные данные).
- `forecast_tables` — извлекает из модели прогнозные значения и формирует таблицу с медианой и 90% доверительным интервалом.
- `build_panel` — строит трёхпанельный график (случаи по регистрации / по заражению / Rt).

Что вы увидите в консоли

```
Вспомогательные функции загружены.
```

ℹ️ Этот блок большой, но выполняется мгновенно: функции не выполняются сами по себе — они просто сохраняются в памяти и вызываются на Шагах 10–11.



15. Шаг 10 — Экспорт национальных результатов

Что делает этот код

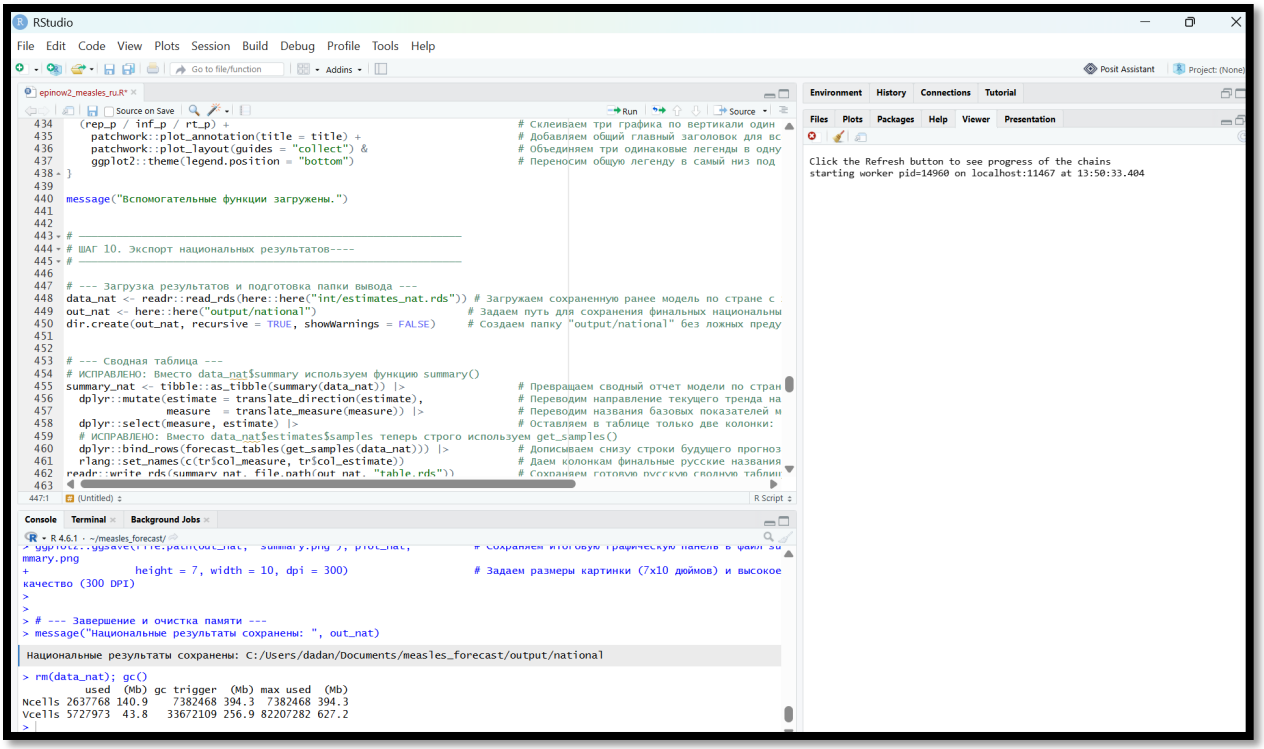
Открывает сохранённый файл `int/estimates_nat.rds` (результат Шага 7) и формирует на его основе результаты в удобном для чтения виде:

- **table.rds** — сводная таблица с оценками Rt, темпов роста, времени удвоения, прогнозного числа случаев (медиана и 90% интервал). Все подписи — на русском языке.
- **summary.png** — трёхпанельный график: верхняя панель — случаи по дате регистрации, средняя — по дате заражения, нижняя — Rt. Заголовок графика — на русском («Кыргызстан»); подписи осей и легенда формируются самим пакетом EpiNow2 и на данный момент остаются на английском (Estimate, Forecast и т.д.) — это известное ограничение текущей версии скрипта, а не ошибка запуска.

Результаты сохраняются в папку `output/national/`.

Что вы увидите в консоли

```
Национальные результаты сохранены: [путь]
```



16. Шаг 11 — Экспорт региональных результатов

Что делает этот код

Открывает `int/estimates_reg.rds` (результат Шага 8) и формирует:

- **summary_plot.png** — сводный график всех регионов на одной картинке (для сравнения Rt и трендов между регионами).
- **summary_table.rds** — сводная таблица по всем регионам (направления трендов, Rt, темпы роста).
- **Для каждого региона** создаётся отдельная папка с summary.png (трёхпанельный график) и table.rds (индивидуальная таблица прогноза).

⚠ summary_plot.png и summary_table.rds читают результат модели через устаревший способ доступа к данным EpiNow2. На части установок (в зависимости от версии пакета) эти два файла могут не создаться — вместо них в консоли появится только предупреждение вида «Предупреждение: сводный график/таблица не создан(а) — ...». Это не останавливает скрипт и не влияет на остальные результаты: таблицы и графики по каждому региону в любом случае будут созданы. Если этих двух файлов нет в output/region/ — это ожидаемое поведение при текущей версии скрипта, а не является ошибкой тренера или студента.

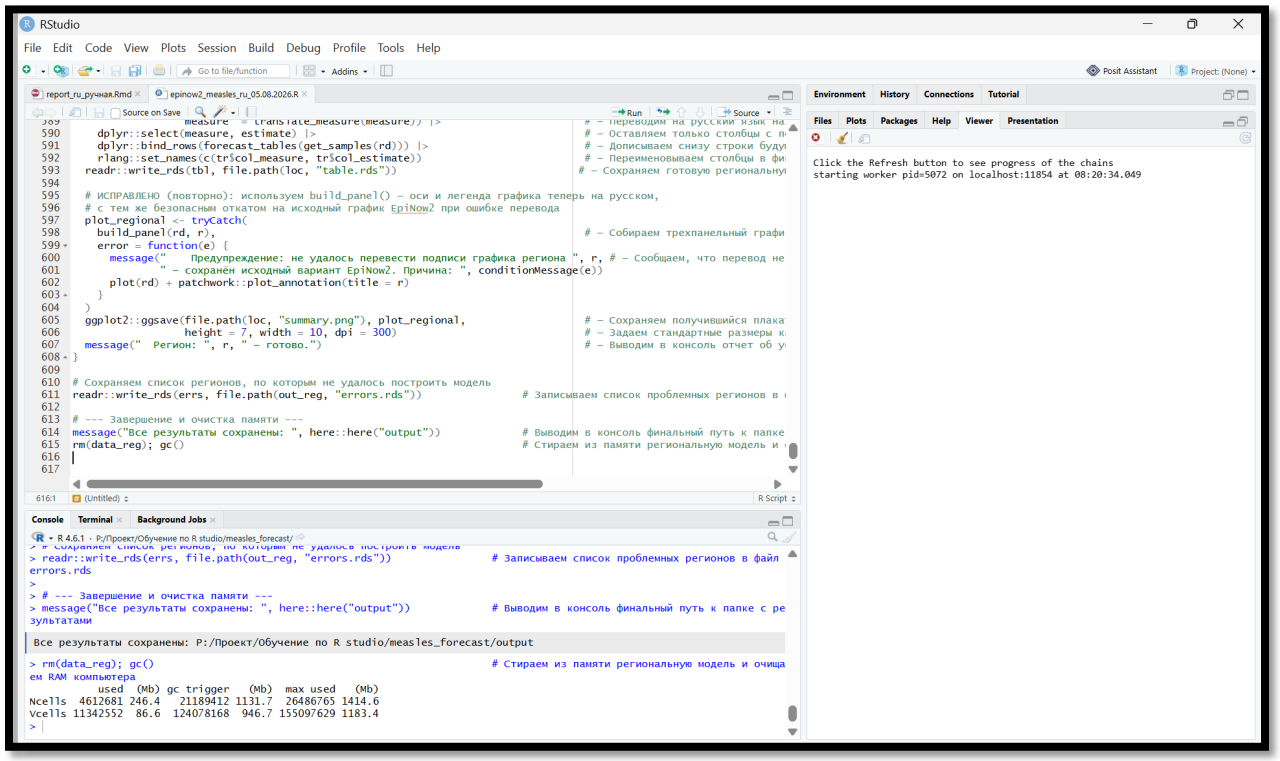
Таблицы (table.rds, summary_table.rds) — полностью на русском языке. У графиков (summary.png по каждому региону) переведён только заголовок (название региона); оси и легенда, как и в национальном графике, на данный момент остаются на английском — это

известное ограничение текущей версии скрипта. Названия папок регионов — в транслитерации (Баткен → Batken) для совместимости с Windows.

Что вы увидите в консоли

Последовательные сообщения по каждому региону:

```
Сводный график регионов сохранён.
Сводная таблица регионов сохранена.
  Регион: г.Бишкек — готово.
  Регион: Чуйская — готово.
  ...
Все результаты сохранены: [путь]
```



ЧАСТЬ III. ЧАСТЫЕ ОШИБКИ И ИТОГОВЫЙ ЧЕК-ЛИСТ

17. Частые ошибки и что делать

Ниже — типичные ситуации, которые возникают в ходе очного обучения, и что с ними делать. Формулировки соответствуют фактическому поведению текущей версии скрипта epinow2_measles_ru.R.

- **«Не найден .xlsx в каталоге data/» на Шаге 4** — папка data/ пуста или скрипт запущен не из R-проекта. Проверьте, что открыт measles_forecast.Rproj (Часть I, п. 4) и что Excel-файл действительно лежит в data/.
- **Ошибка чтения Excel ещё до появления сообщения в консоли** — в data/ больше одного .xlsx-файла. Оставьте только один нужный файл (см. предупреждение в п. 3.1).
- **«0 случаев» или явно заниженное число случаев** — номера столбцов в Шаге 2 (COL_REGION, COL_REGDATE, COL_CONFIRM) не совпадают со структурой вашего Excel, либо названия статусов в файле отличаются от CONFIRM_OK.
- **Число регионов не совпадает с ожидаемым (например, 9 вместо 8)** — почти всегда опечатка или лишний пробел в названии региона в Excel. Сверьте уникальные названия региона в данных с восемью названиями регионов в ref.csv и словаре tr$region_names (Шаг 9).
- **Предупреждение «Национальная задержка применена к регионам без записи в ref.csv»** — для одного или нескольких регионов не найдена строка в справочнике — обычно из-за расхождения в написании названия. Это не останавливает расчёт: региону просто присваивается национальная задержка.
- **Шаг 7 или Шаг 8 выполняется очень долго или кажется «зависшим»** — это нормально: выполнение национальной модели занимает 30–60 минут, региональной — 30–90 минут. Признак того, что расчёт идёт: в консоли обновляется прогресс сэмплирования (номер цепи, % завершения). Не закрывайте RStudio и не выключайте компьютер. Планируйте перерыв именно между Шагом 7 и Шагом 8.
- **В output/region/ нет summary_plot.png или summary_table.rds** — в текущей версии скрипта эти два сводных файла читаются через устаревший способ доступа к результатам модели и на части установок EpiNow2 могут не создаваться — вместо создания файла в консоли появится только предупреждение. Это не затрагивает остальные результаты (таблицы и графики по каждому региону).
- **На графике summary.png оси и легенда на английском** — заголовок графика (страна или регион) переведён на русский, но подписи осей и легенда формируются самим пакетом EpiNow2 и на данный момент остаются на английском. Это известное ограничение текущей версии скрипта, не ошибка запуска.

ℹ Если ни один пункт не подошёл — запишите точный текст ошибки из консоли и номер шага, на котором она возникла: это сильно ускорит разбор проблемы.

18. Итоговый чек-лист: что должно получиться

После успешного выполнения всех 11 шагов в папке проекта должны появиться следующие файлы. Используйте список, чтобы быстро проверить результат работы каждого студента.

Кеш моделей (папка int/)

- int/estimates_nat.rds — результат Шага 7 (национальная модель)
- int/estimates_reg.rds — результат Шага 8 (региональная модель)

Национальные результаты (папка output/national/)

- table.rds — сводная таблица (полностью на русском)
- summary.png — трёхпанельный график (заголовок на русском, оси/легенда — на английском)

Региональные результаты (папка output/region/)

- summary_table.rds и summary_plot.png — сводные по всем регионам (могут отсутствовать, см. п. 17)
- region_map.rds и errors.rds — служебные файлы соответствия папок и списка проблемных регионов
- по одной подпапке на каждый из 8 регионов, в каждой — table.rds и summary.png

ℹ Если какого-то файла нет — сначала сверьтесь с разделом 17 «Частые ошибки», прежде чем считать это сбоем.