

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ОБРАЗОВАНИЯ**

**Национальный исследовательский ядерный университет
«МИФИ»**

Кафедра №71

Факультет Бизнес-информатики и управления
комплексными системами (ФБИУКС)

Экзаменационная работа на тему:

«Исследование взаимосвязи популярности интернет-запросов и числа
новых случаев заражения коронавирусом»

Подготовил: студент учебной группы Б19-902

Шопхоев Д. С.

Проверяющий: доцент кафедры региональной
и инновационной экономики

Смирнов Д. С.

Москва 2022г.

Оглавление

Оглавление	2
Введение.....	3
Материалы и методы	4
Результаты исследования	7
Россия	7
Франция.....	26
США	42
Индия.....	58
Великобритания.....	74
ЮАР.....	90
Бразилия	106
Исследование корреляции ключевых слов по странам:.....	122
Заключение	124
Список литературы	125

Введение

Пандемия коронавирусной инфекции (COVID-19) создала уникальную возможность установить закономерности активности пользователей в интернете в связи с новым заболеванием.

Цель исследования. Оценка взаимосвязи интереса пользователей к интернет-запросам, связанным с COVID-19, в странах, в которых фиксировали первые случаи заражения штаммами, вызывающими беспокойство, а именно [1]:

Штамм	Дата первого зарегистрированного случая	Страна
Alpha	Sep-2020	United Kingdom
Beta	May-2020	South Africa
Gamma	Nov-2020	Brazil
Delta	Oct-2020	India
Omicron	Nov-2021	South Africa

Таблица 1 Штаммы вызывающие беспокойство

Также были проанализированы следующие страны: Россия, США, Франция. Оценка динамики эпидемического процесса этой инфекции по поисковым запросам в Google Trends с целью возможного контроля заболеваемостью и своевременного введения ограничений. Сравнение результатов более ранних аналогичных исследований.

Материалы и методы

Базы данных случаев заболевания

Данные по ежедневно подтвержденным случаям заболевания COVID-19 в России, Франции, США, Великобритании, Бразилии и ЮАР были получены с использованием ресурса www.ourworldindata.org/covid-cases [2].

Базы данных поисковых запросов

Google Trends [3] является общедоступной базой данных выборки реальных поисковых запросов, выполненных с использованием поисковой системы Google, которые анонимны, классифицированы и агрегированы. Данные представляют из себя временной ряд значений, представленных в интервале от 0 до 100. Google Trends не позволяет получить абсолютные значения количества запросов, что несколько затрудняет исследование.

Проведен анализ запросов по теме «COVID-19», полученных из Google Trends, с 30.01.2020 по 14.01.2022.

Выбор терминов по номенклатуре и симптомам заболевания

Предварительный список поисковых терминов был составлен в соответствии с симптоматикой COVID-19 [4]. В проведенной работе оценивались запросы пользователей из следующих стран: Россия, Франция, США, Великобритания, Бразилия и ЮАР на официальном языке страны. В перечень поисковых запросов, обозначающих симптомы заболевания, вошли «боль в горле», «заложенность носа», «кашель», «лихорадка», «обоняние», «одышка», «температура», «головная боль». Помимо симптомов в список поисковых запросов включены все запросы по теме COVID-19 (алгоритмы Google позволяют агрегировать конкретные запросы в группы по темам) Исследуемый период был установлен с 30.01.2020 по 14.01.2022.

Методы При сравнении данных GT для каждого симптома и термина со случаями COVID-19 использовали коэффициент корреляции Пирсона (r), коэффициент корреляции Спирмена (R), а также кросскорреляцию.

Коэффициент корреляции Пирсона

Коэффициент корреляции Пирсона характеризует существование линейной зависимости между двумя величинами.

Пусть даны две выборки:

$$x^m = (x_1, \dots, x_m), \quad y^m = (y_1, \dots, y_m);$$

коэффициент корреляции Пирсона рассчитывается по формуле:

$$r_{xy} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \sum_{i=1}^m (y_i - \bar{y})^2}} = \frac{\text{cov}(x, y)}{\sqrt{s_x^2 s_y^2}},$$

где \bar{x}, \bar{y} – выборочные средние x^m и y^m , s_x^2, s_y^2 – выборочные дисперсии, $r_{xy} \in [-1, 1]$.

Коэффициент корреляции Пирсона называют также теснотой линейной связи.

Коэффициент корреляции Спирмена

Заданы две выборки $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$.

Коэффициент корреляции Спирмена вычисляется по формуле:

$$\rho = 1 - \frac{6}{n(n-1)(n+1)} \sum_{i=1}^n (R_i - S_i)^2,$$

где R_i – ранг наблюдения x_i в ряду x , S_i – ранг наблюдения y_i в ряду y .

Коэффициент ρ принимает значения из отрезка $[-1; 1]$.

Равенство $\rho = 1$ указывает на строгую прямую линейную зависимость, $\rho = -1$ на обратную.

Кросскорреляция (взаимнокорреляционная функция) — стандартный метод оценки степени корреляции двух последовательностей. Она часто используется для поиска в длинной последовательности более короткой заранее известной. Рассмотрим два ряда f и g . Взаимная корреляция определяется по формуле:

$$(f \star g)_i \stackrel{\text{def}}{=} \sum_j f_j^* g_{i+j}$$

где i — сдвиг между последовательностями относительно друг друга, а верхний индекс в виде звёздочки означает комплексное сопряжение.

Результаты исследования

Россия

Таблица 2 Корреляция заболеваемости и популярности запросов в России

WORDKEY	PEARSON	SPEARMAN
COVID	0.329257	0.395851
SMELL	0.396693	0.55849
THROAT	-0.11343	-0.13235
FEWER	0.17487	0.315999
DYSPNEA	0.129586	0.183216
COUGH	0.340344	0.283081
RHEUM	0.253952	0.256227
HEADACHE	0.047515	0.102577

Таблица 2 содержит информацию о коэффициентах корреляции Спирмена и Пирсона популярности запросов и заболеваемости в России. В таблице приведены несдвинутые временные ряды, поэтому корреляция может быть не максимальна. Для более глубокого анализа запроса в выбранной стране предлагается рассмотреть график кросскорреляционной функции.

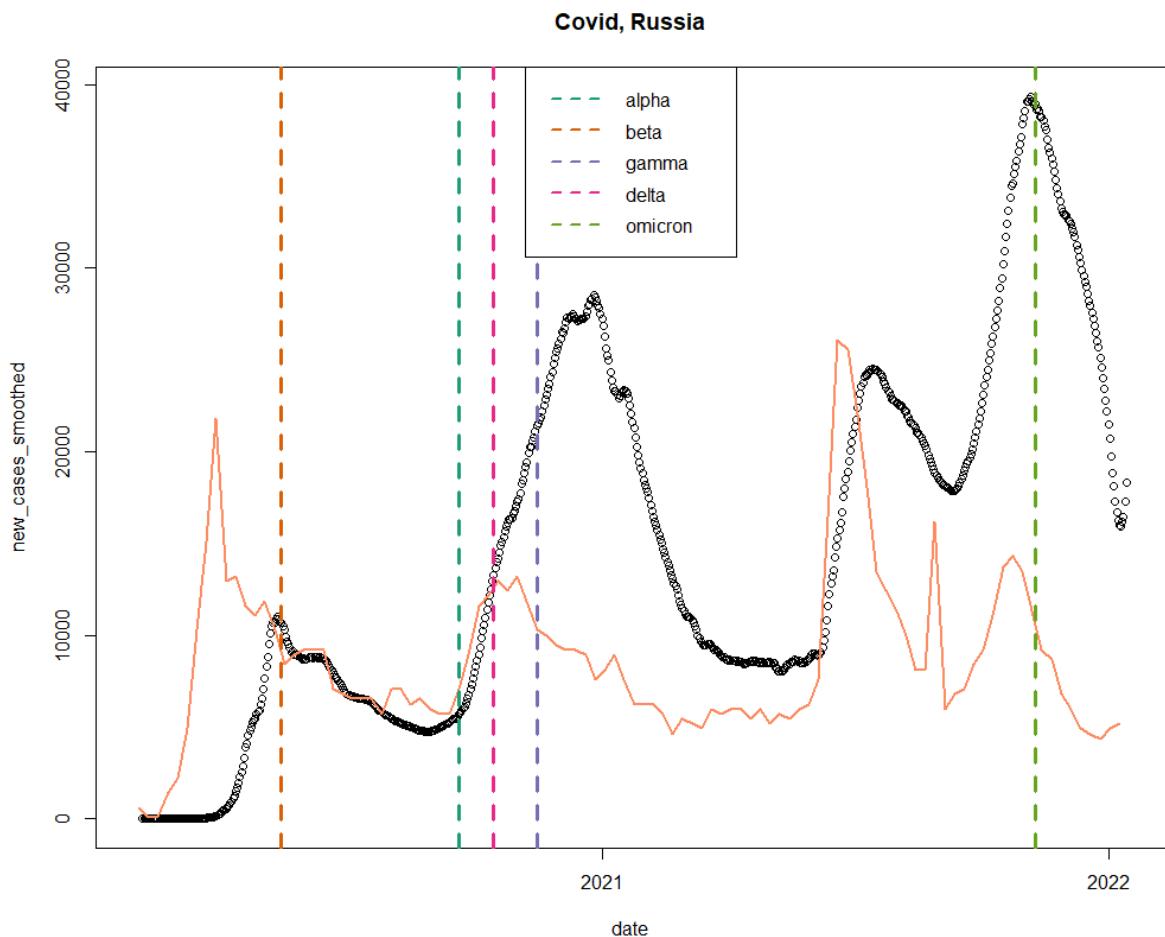


Рисунок 1 Графики популярности запросов по теме COVID-19 и подтвержденных случаев заболевания в России

График популярности запросов по теме COVID-19 окрашен в оранжевый, подтвержденные случаи заболевания — в черный. Вертикальными линиями отмечены даты первых случаев заражения указанным штаммом (Таблица 1).

Важно отметить, что график популярности отмасштабирован для удобства восприятия. Значения популярности запроса ограничены от 0 до 100.

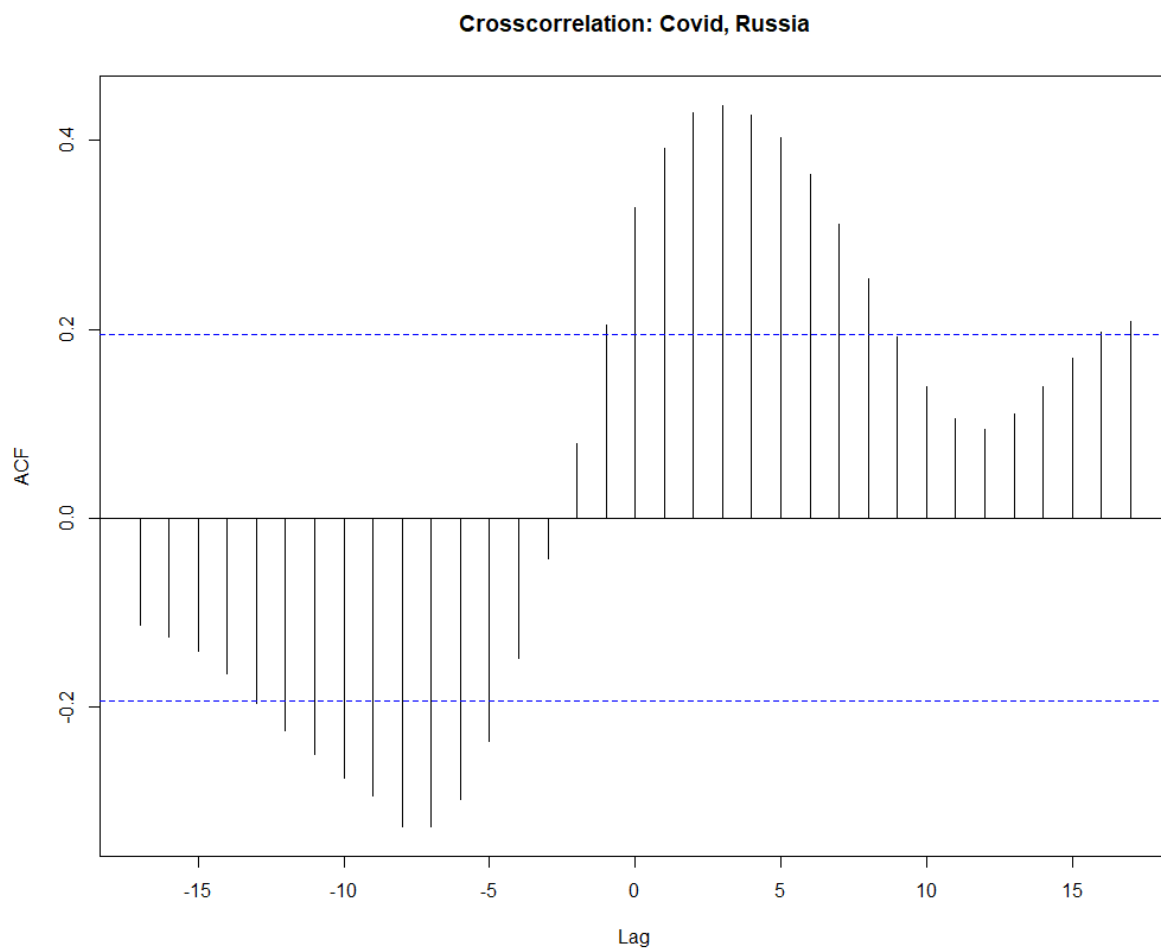


Рисунок 2 График кросскорреляции запросов по теме COVID-19 и заболеваемостью в России

График кросскорреляционной функции показывает корреляцию в зависимости от сдвига (лага). На данном графике хорошо виден лаг: максимум корреляции достигается при $\text{Lag} = 4$ (1 $\text{Lag} = 1$ неделя).

Уровень значимости для коэффициента корреляции $\hat{\rho} \sim 0.2$ при $p = 0.95$, $n = 100$

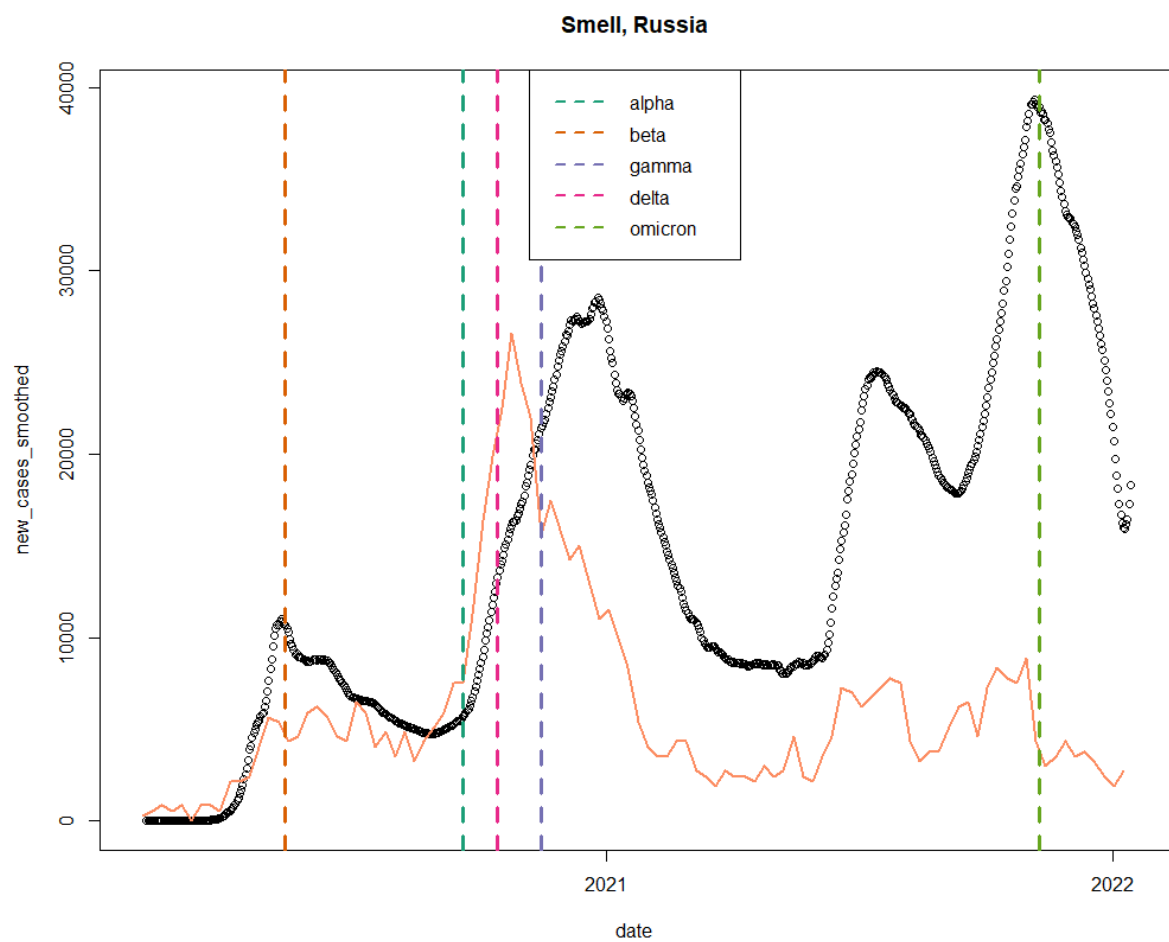


Рисунок 3 Графики популярности запросов по теме Обоняние и подтвержденных случаев заболевания в России

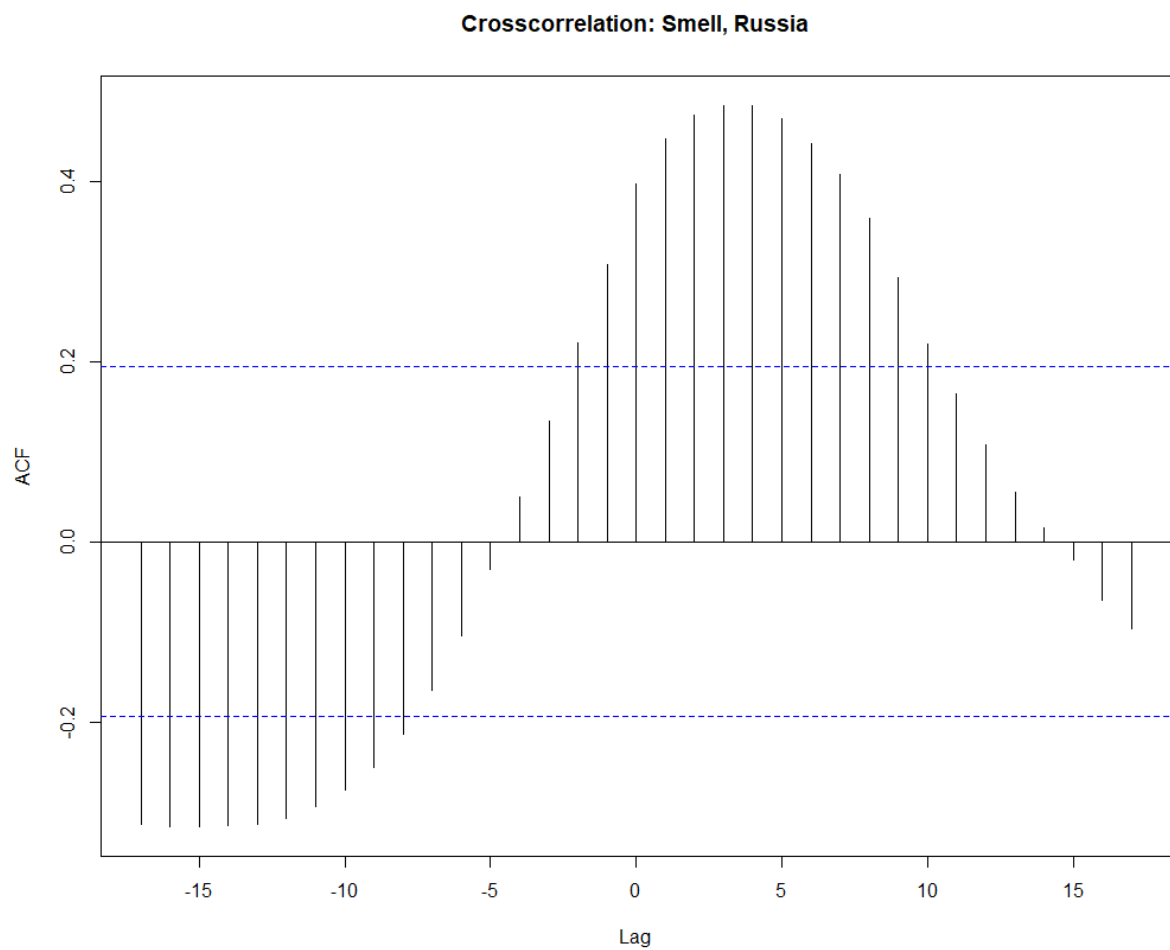


Рисунок 4 График кросскорреляции запросов по теме Обоняние и заболеваемостью в России

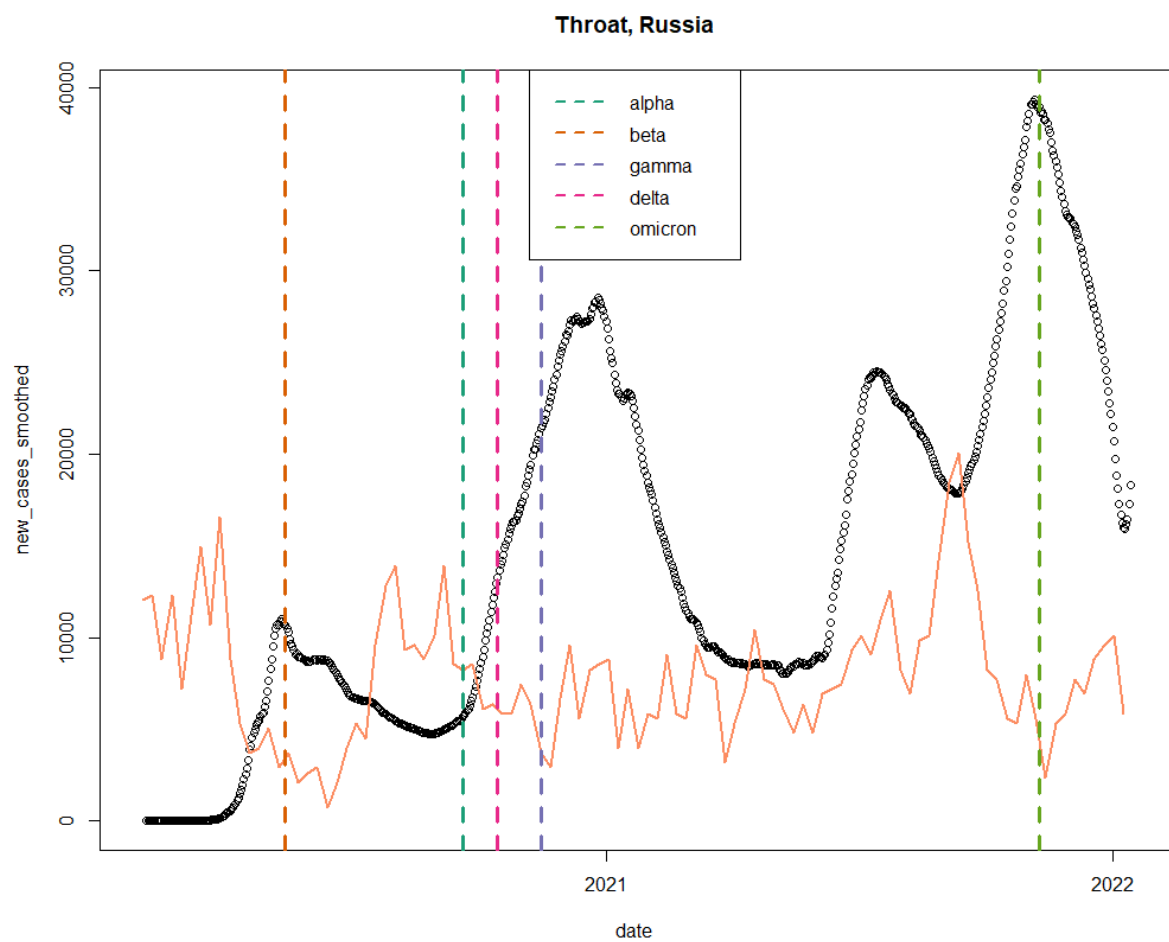


Рисунок 5 Графики популярности запросов по теме Боль в горле и подтвержденных случаев заболевания в России

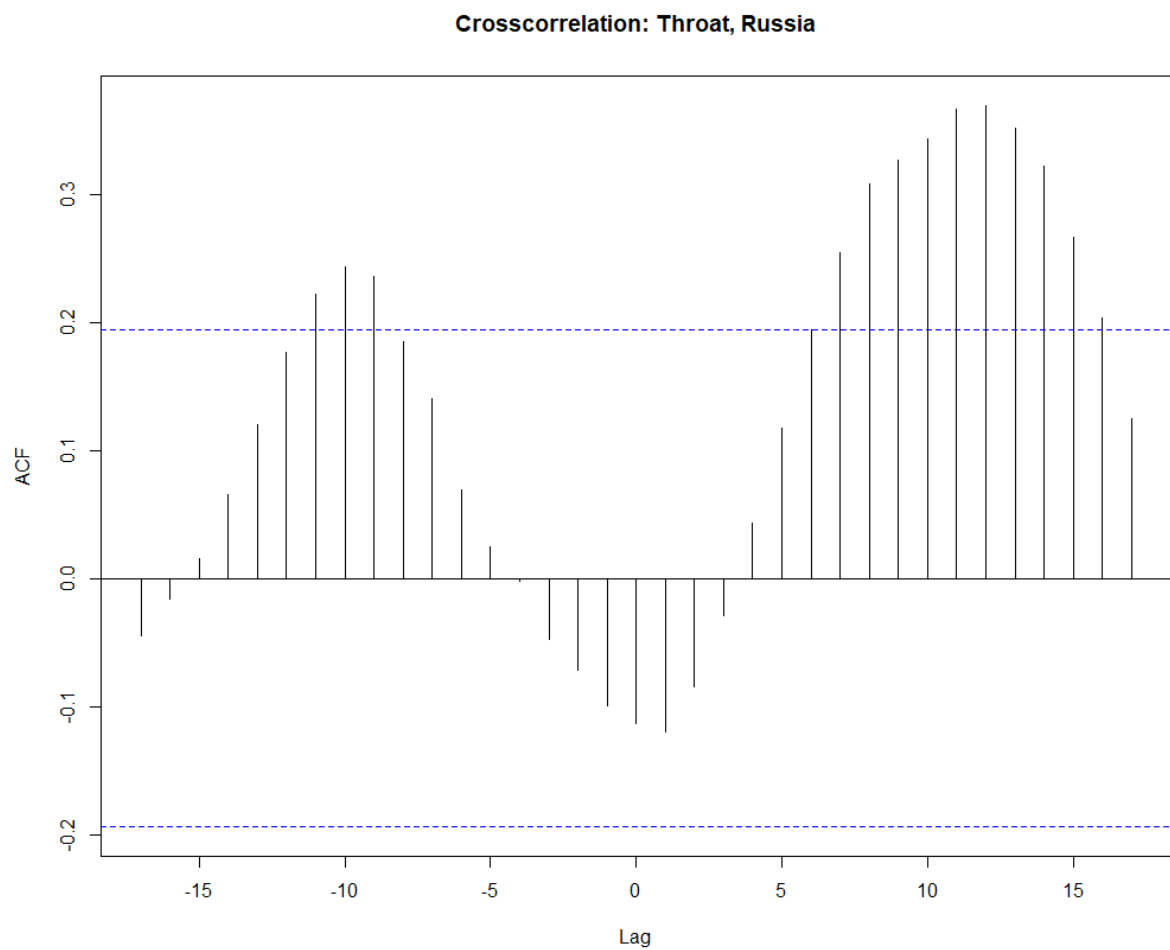


Рисунок 6 График кросскорреляции запросов по теме Боль в горле и заболеваемостью в России

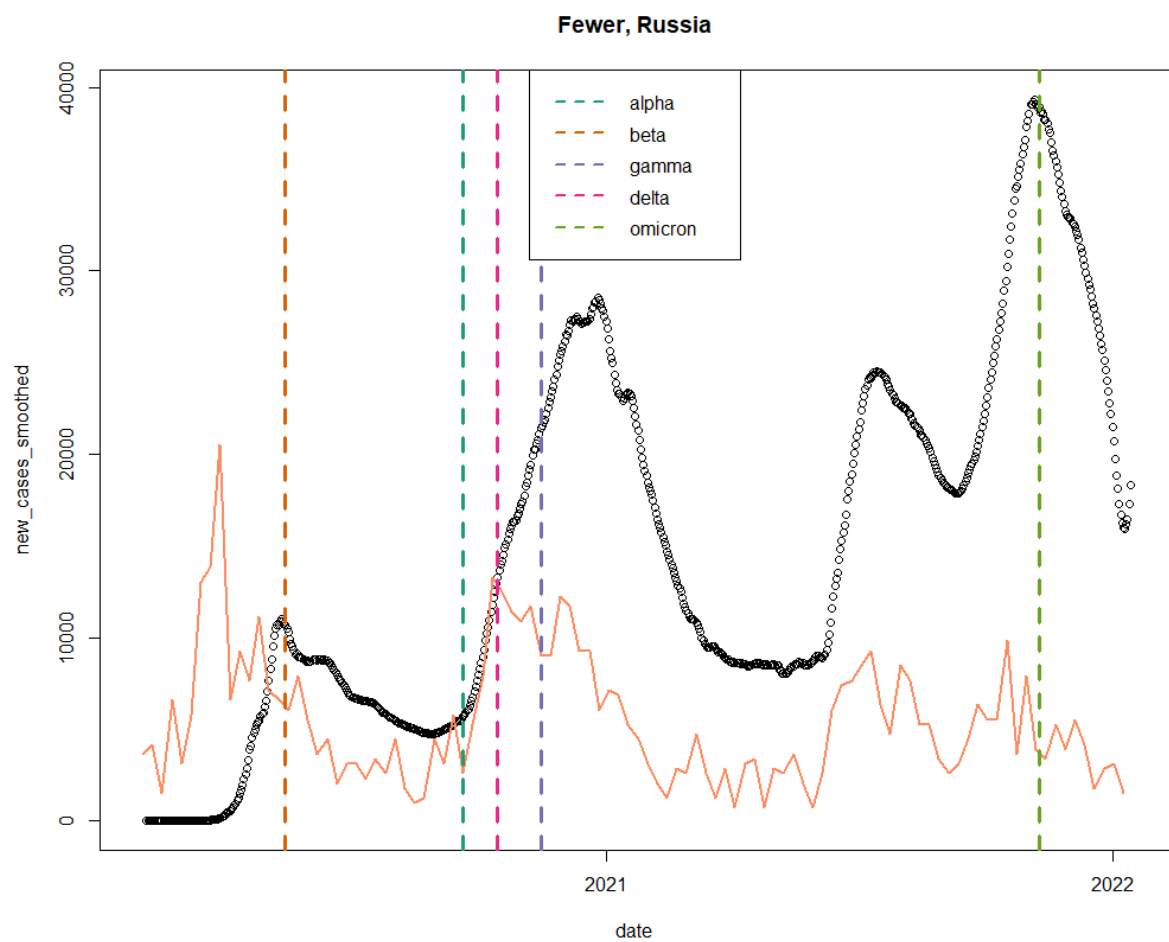


Рисунок 7 Графики популярности запросов по теме Лихорадка и подтвержденных случаев заболевания в России

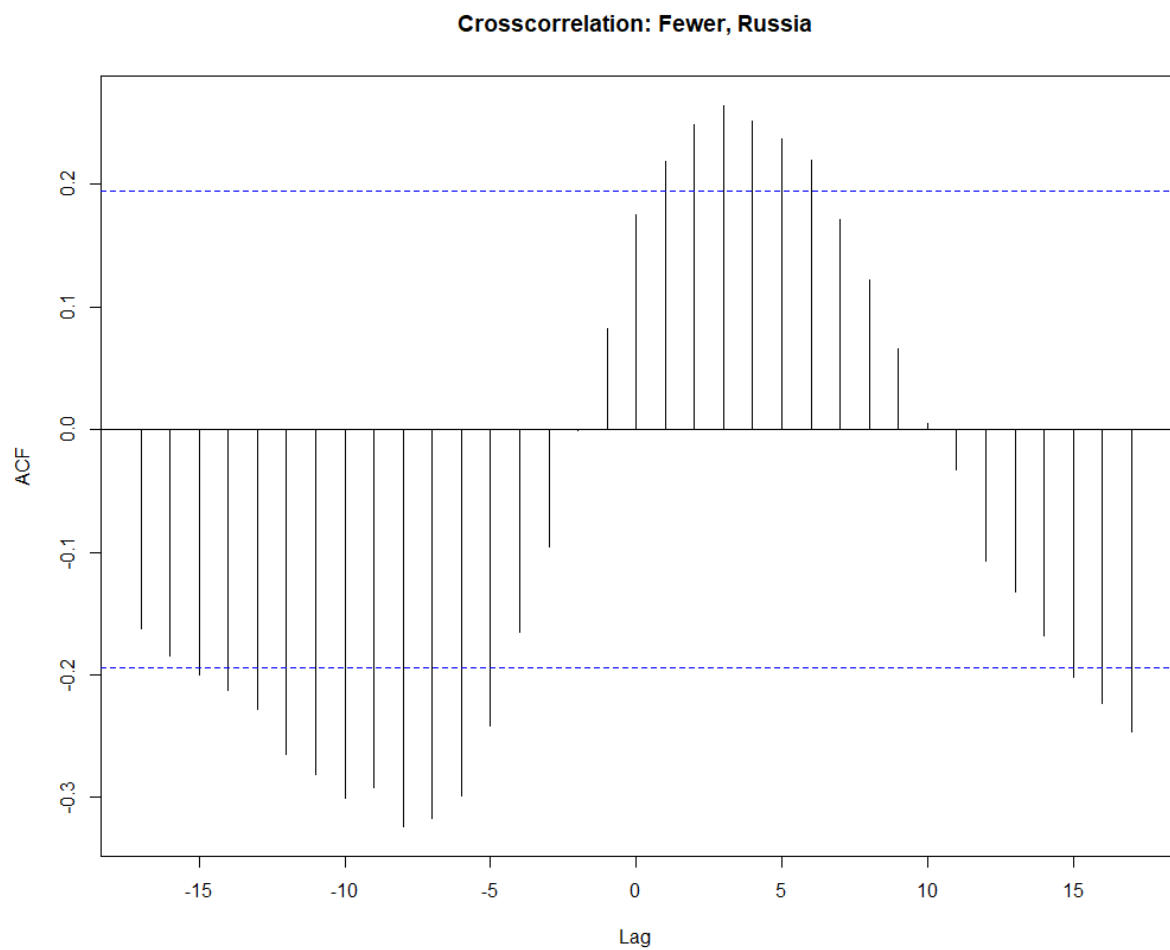


Рисунок 8 График кросскорреляции запросов по теме Лихорадка и заболеваемостью в России

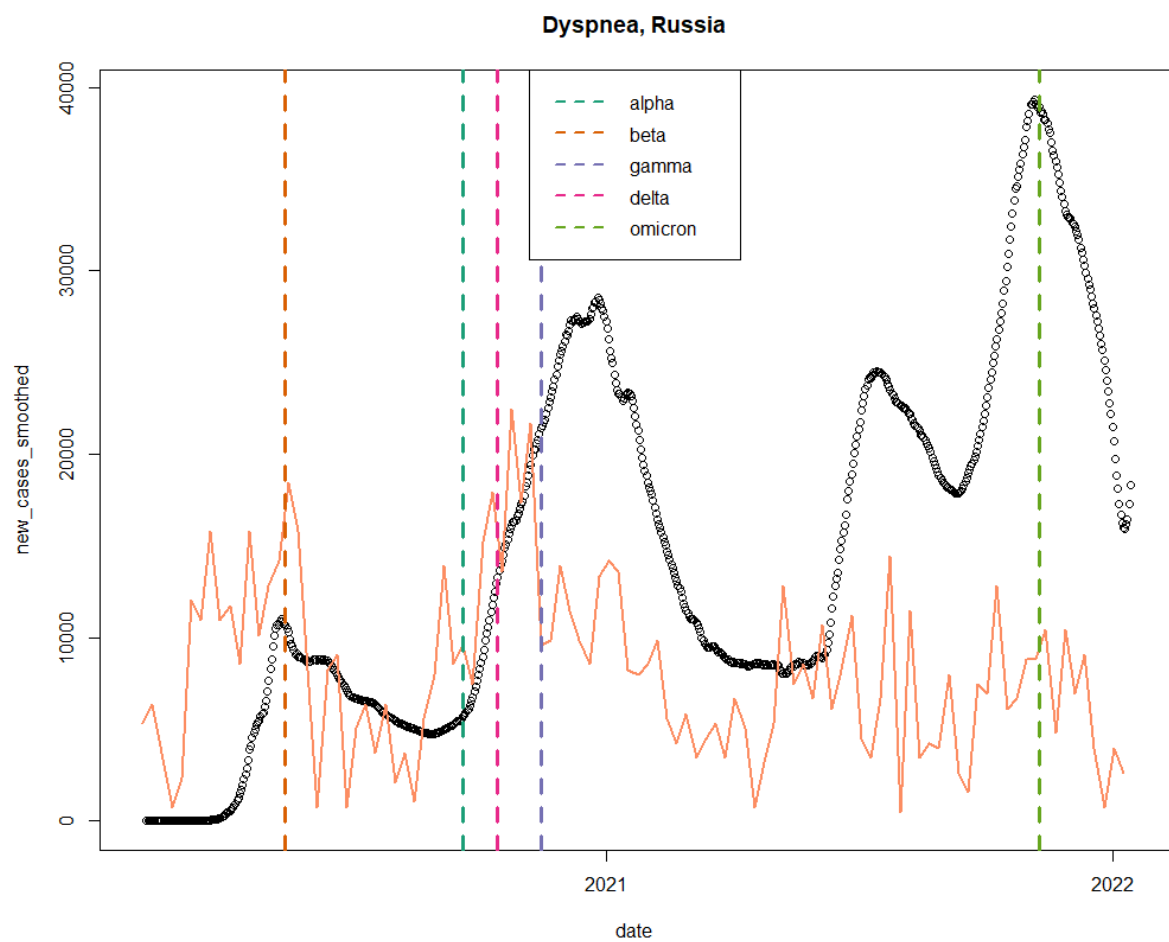


Рисунок 9 Графики популярности запросов по теме Одышка и подтвержденных случаев заболевания в России

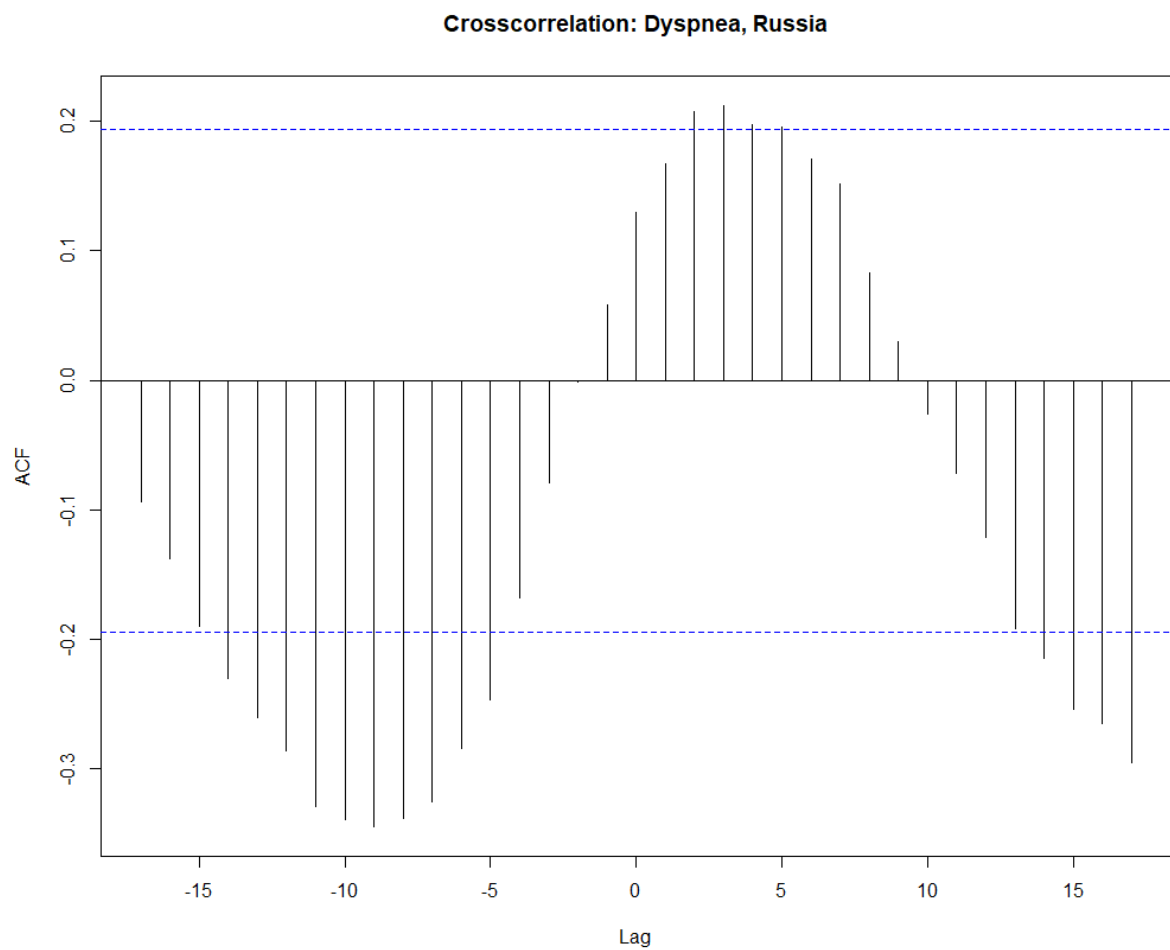


Рисунок 10 График кросскорреляции запросов по теме Одышка и заболеваемостью в России

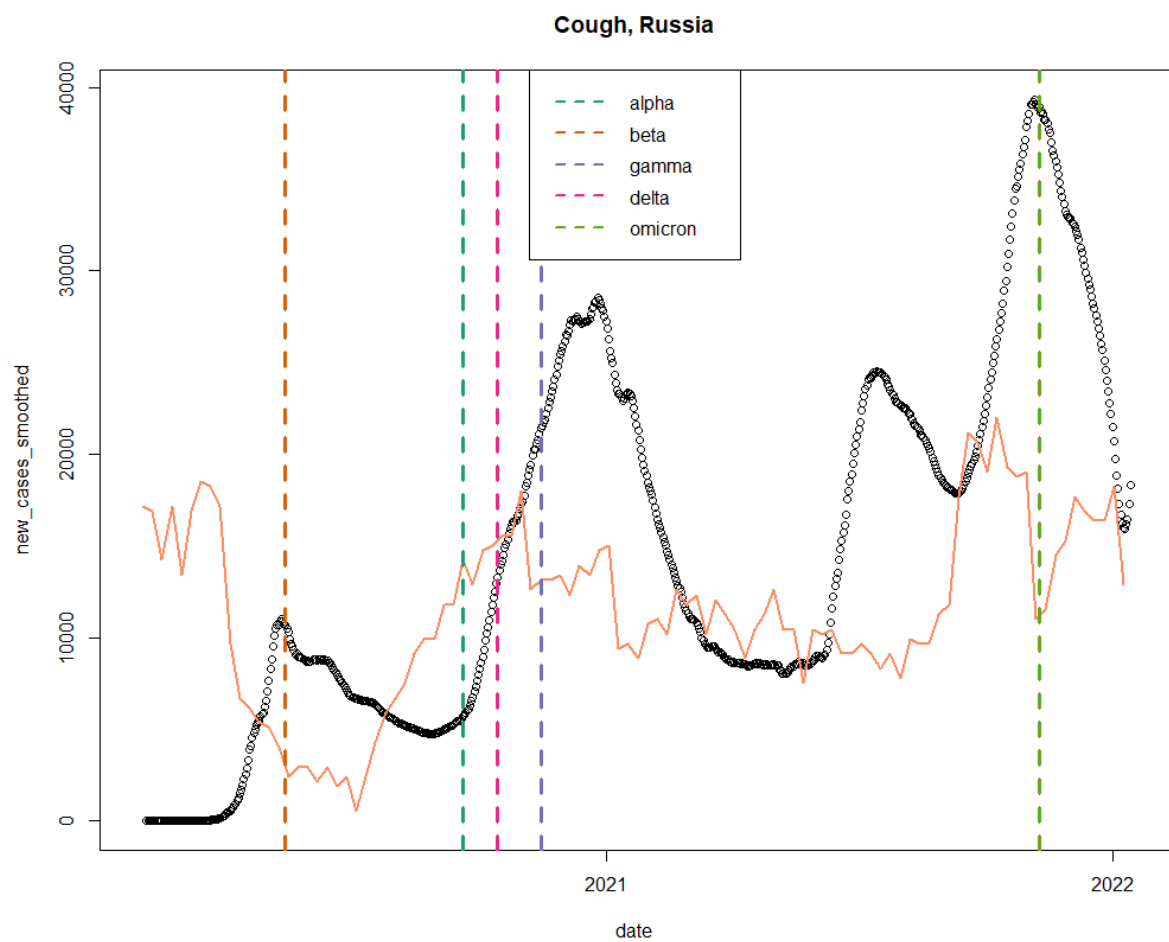


Рисунок 11 Графики популярности запросов по теме Кашель и подтвержденных случаев заболевания в России

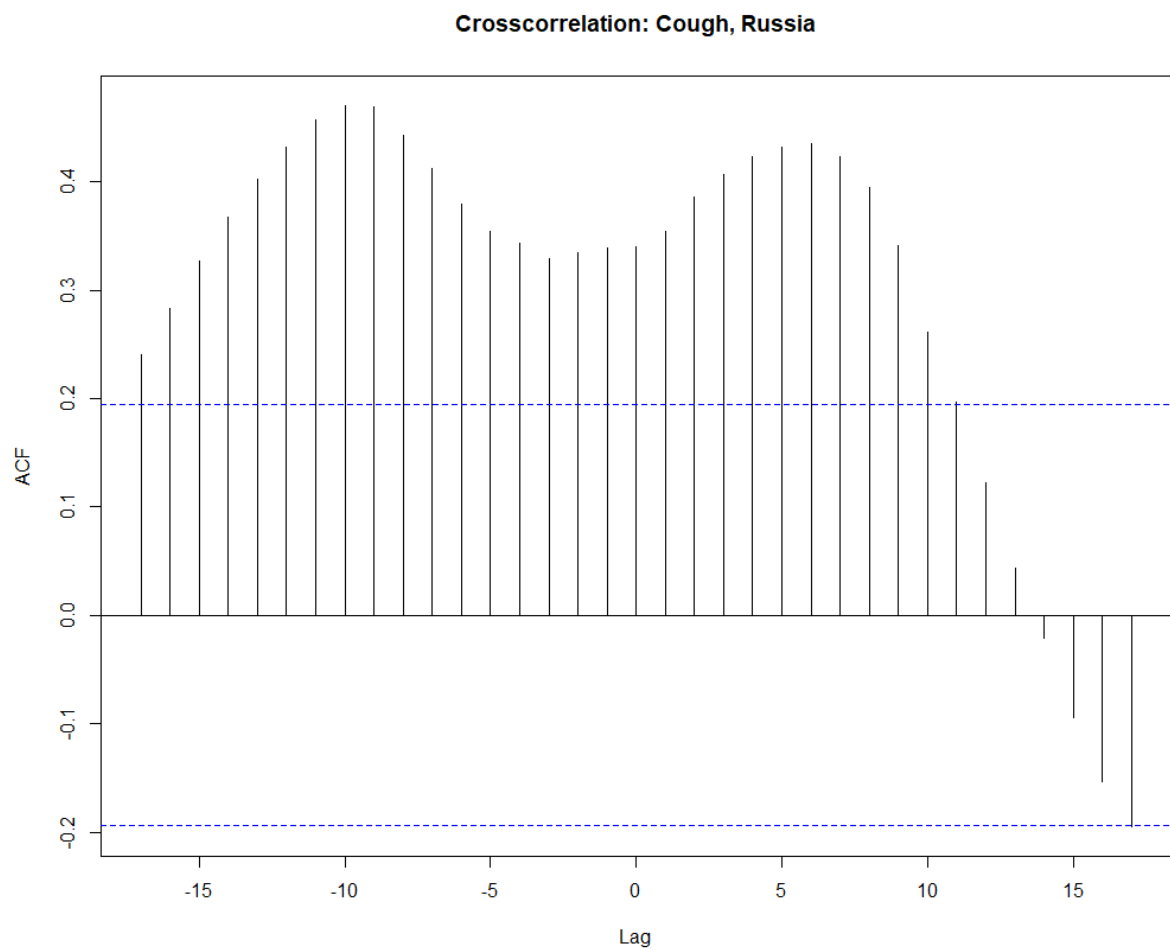


Рисунок 12 График кросскорреляции запросов по теме Кашель и заболеваемостью в России

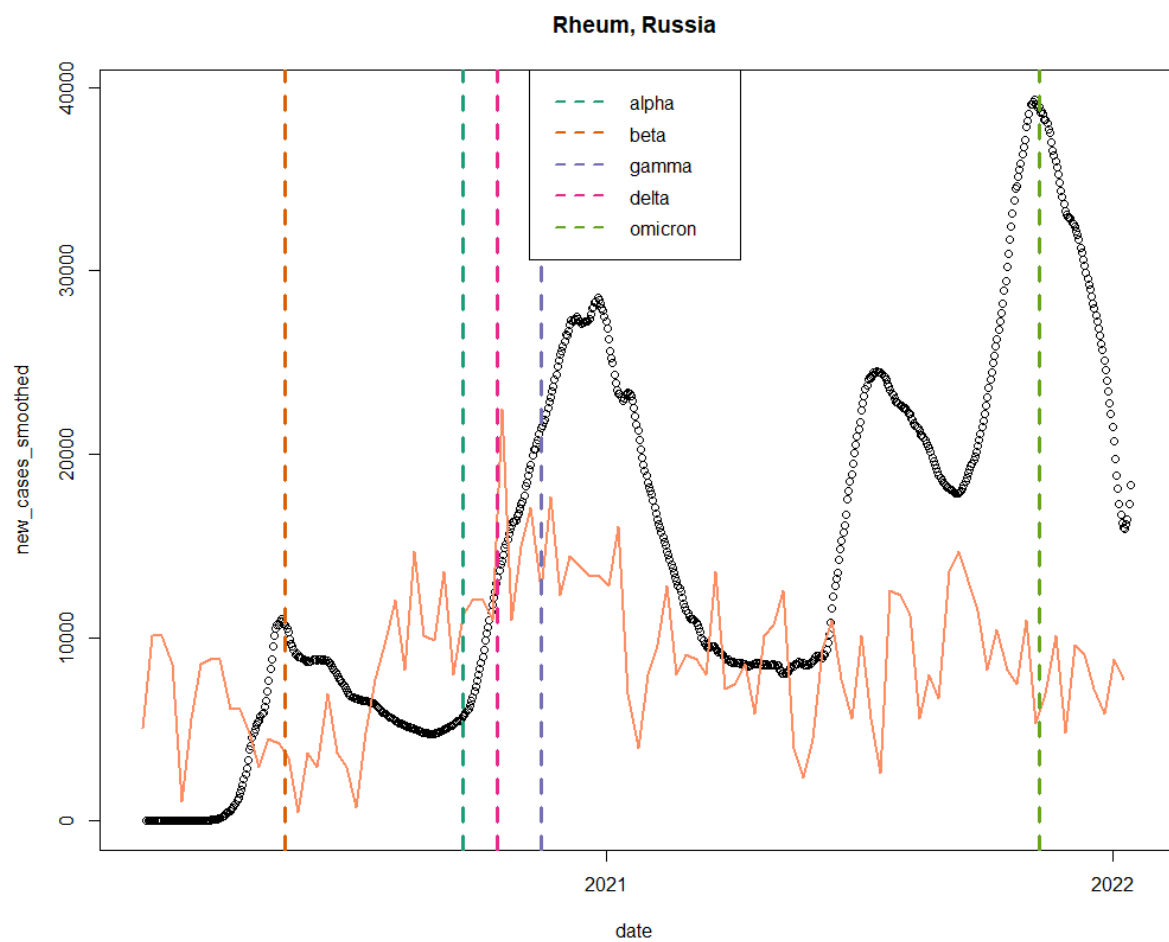


Рисунок 13 Графики популярности запросов по теме Насморк и подтвержденных случаев заболевания в России

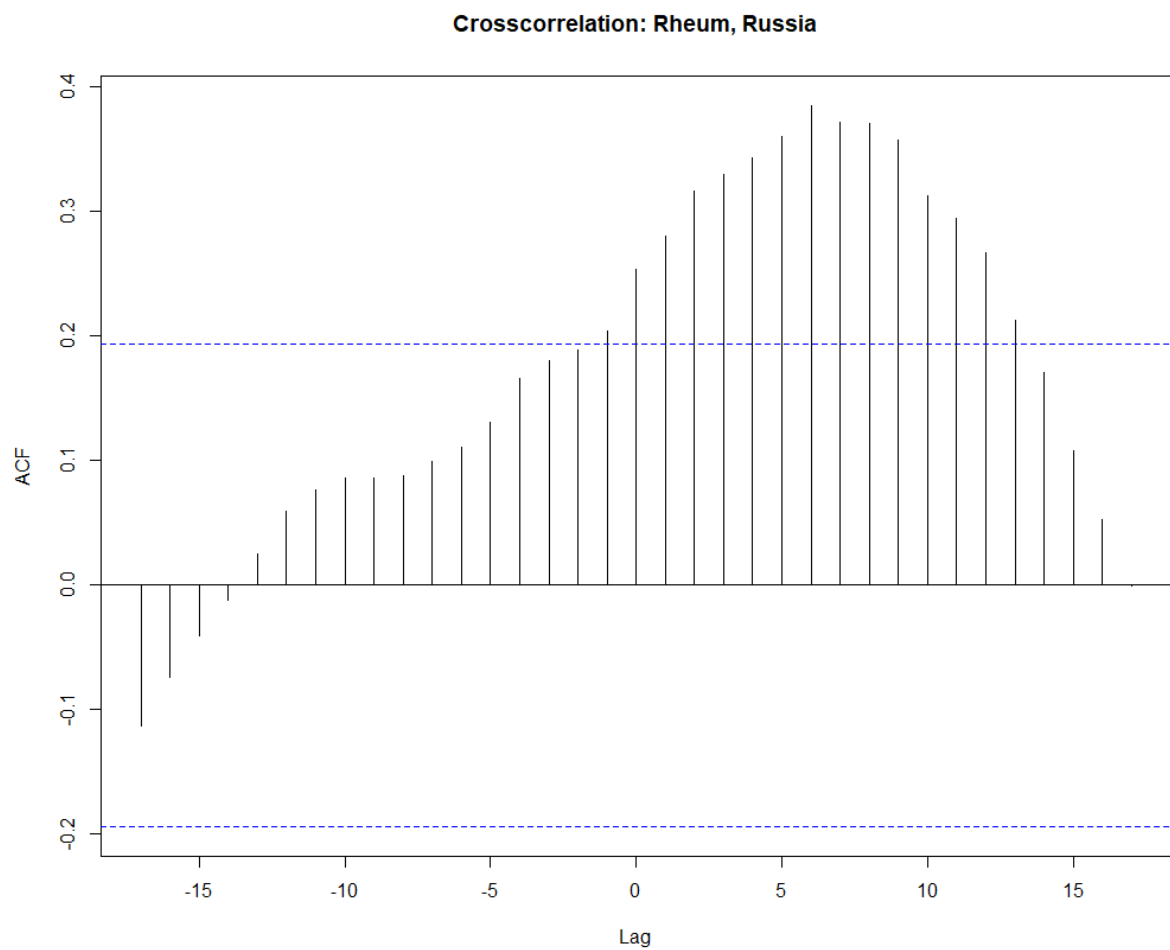


Рисунок 14 График кросскорреляции запросов по теме Насморк и заболеваемостью в России

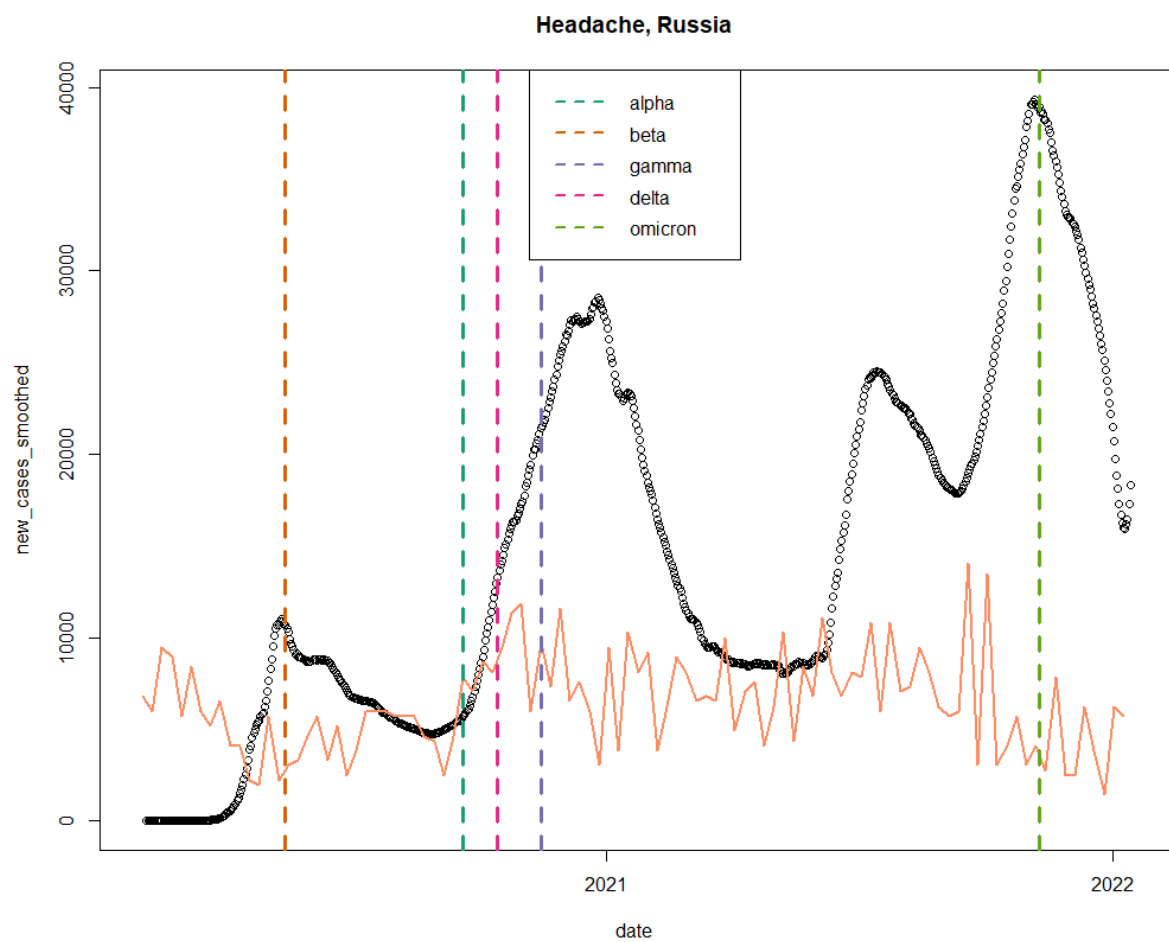


Рисунок 15 Графики популярности запросов по теме Головная боль и подтвержденных случаев заболевания в России

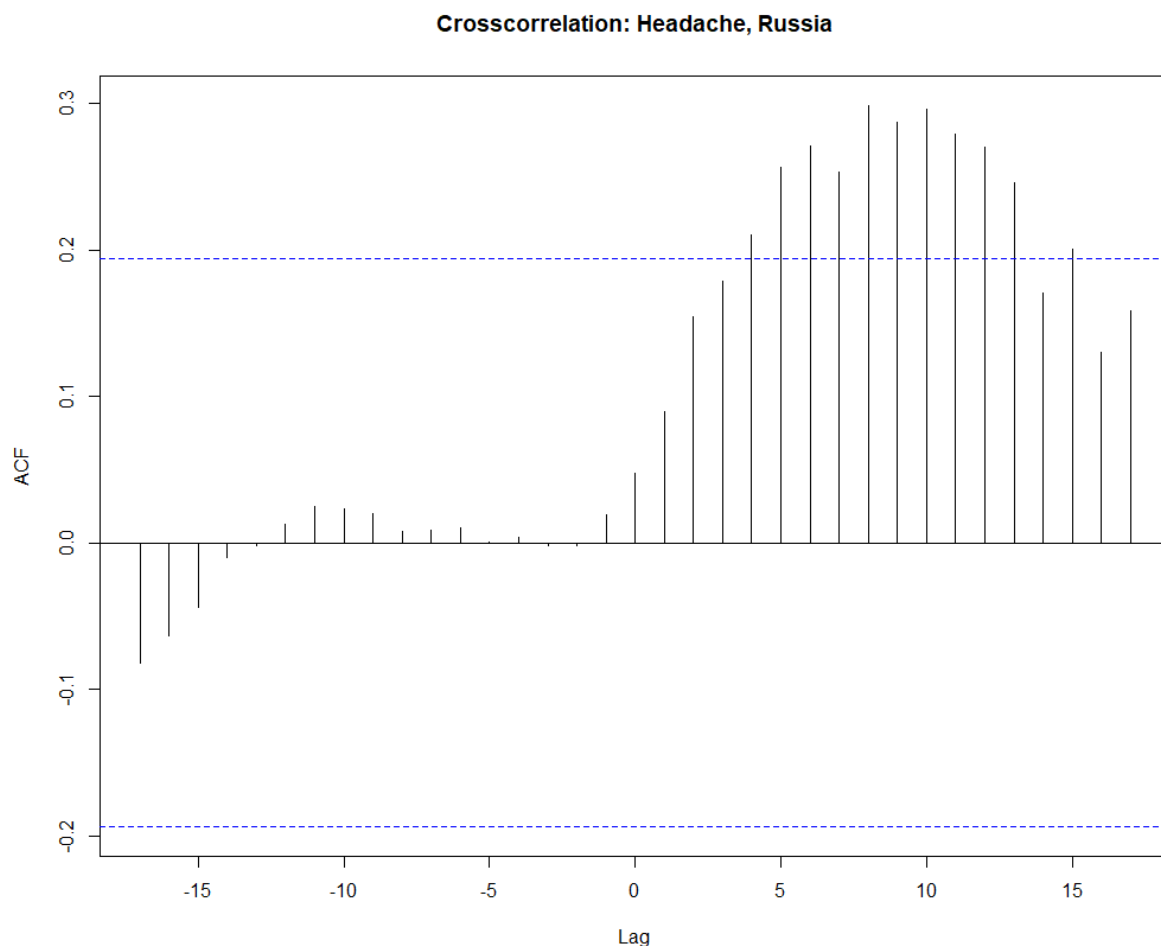


Рисунок 16 График кросскорреляции запросов по теме Головная боль и заболеваемостью в России

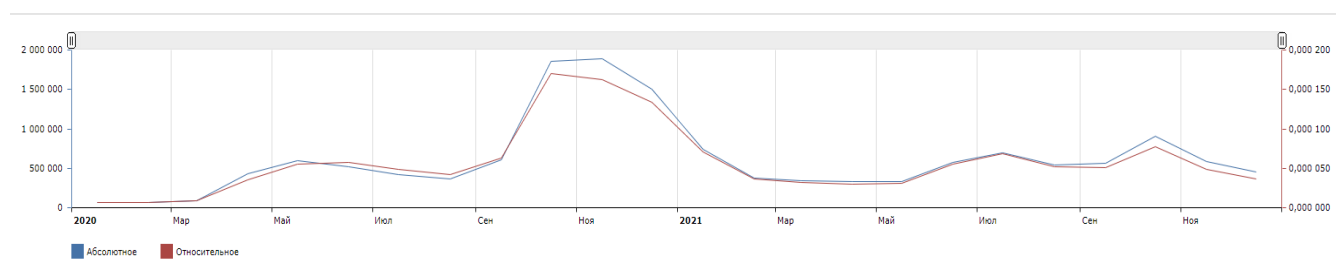
Анализ данных по стране:

Россия на большинстве показывает максимум функции при лаге, в среднем 4-5. Это может говорить о том, что официальная статистика несколько отстает от реальных заражений, причем отставание в два раза больше, чем инкубационный период коронавируса, который составляет до 14 дней [5].

Исследование 2020 года показывает отрицательную корреляцию с довольно большим модулем для динамики заболеваемости и популярностью запроса по теме «Кашель» [6]. Если мы посмотрим на график, то можем заметить резкий спад при росте в первую волну. Сейчас ситуация изменилась: корреляция $r=0.34$ Можно предположить, что это также связано с лагом.

Однако если мы посмотрим на график популярности запроса по теме «Обоняние», как на самый коррелируемый, то можем увидеть тот самый лаг = 4 на второй волне, но на остальных такого сильного отставания не наблюдается.

Для России есть еще один альтернативный источник данных: Яндекс.Подбор слов. Это аналогичный GoogleTrends сервис, позволяющий изучать историю показов по фразе. Здесь намного меньшая дискретизация, но он позволяет видеть не относительные значения (от 0 до 100), как в GoogleTrends, а абсолютное число показов. К сожалению, включение в исследование данных из этого сервиса затруднительно из-за отсутствия функционала для скачивания данных и капчи при каждом запросе. Приведем только пример для фразы «Обоняние»



01.02.2020 - 29.02.2020	62553	0,000006688625
01.03.2020 - 31.03.2020	85790	0,000008585426
01.04.2020 - 30.04.2020	433128	0,000035695906
01.05.2020 - 31.05.2020	597426	0,000054790254
01.06.2020 - 30.06.2020	517747	0,000057169917
01.07.2020 - 31.07.2020	416258	0,000048696040
01.08.2020 - 31.08.2020	359175	0,000042203308
01.09.2020 - 30.09.2020	603066	0,000062579224
01.10.2020 - 31.10.2020	1855419	0,000169842223
01.11.2020 - 30.11.2020	1893034	0,000162568920
01.12.2020 - 31.12.2020	1497528	0,000133482708
01.01.2021 - 31.01.2021	745334	0,000070296536
01.02.2021 - 28.02.2021	373818	0,000036834192
01.03.2021 - 31.03.2021	347940	0,000031500956
01.04.2021 - 30.04.2021	331186	0,000030267853
01.05.2021 - 31.05.2021	332587	0,000030807440
01.06.2021 - 30.06.2021	572538	0,000055001125
01.07.2021 - 31.07.2021	698057	0,000068452054
01.08.2021 - 31.08.2021	544195	0,000052047391

01.09.2021 - 30.09.2021	558484	0,000050674126
01.10.2021 - 31.10.2021	906314	0,000077384605
01.11.2021 - 30.11.2021	587786	0,000049084990
01.12.2021 - 31.12.2021	447733	0,000036417578

В месяца пика второй волны (октябрь-ноябрь 2020) было свыше 1,8 млн запросов в каждом из месяцев. В то же время по официальной статистике было зарегистрировано около миллиона случаев заражения в каждый из двух месяцев.

По данным сайта доля Google в России в 2020 составляет 60,3% против 38,1% у Яндекса. Таким образом мы можем получить очень грубую оценку абсолютного числа запросов по теме «Обоняние» в России во вторую волну:

$1800000 / 0.381 = 4720000$ – за один месяц (октябрь-ноябрь)

Это также косвенно может дать оценку реальному количеству заболеваний.

Франция

Таблица 3 Корреляция заболеваемости и популярности запросов во Франции

WORDKEY	PEARSON	SPEARMAN
COVID	0.469812	0.779743
SMELL	0.285619	0.638629
THROAT	0.115282	0.058643
FEWER	0.084024	0.185965
DYSPNEA	-0.03531	0.012964
COUGH	0.150052	0.170799
RHEUM	0.209374	0.21905
HEADACHE	0.071745	0.183587

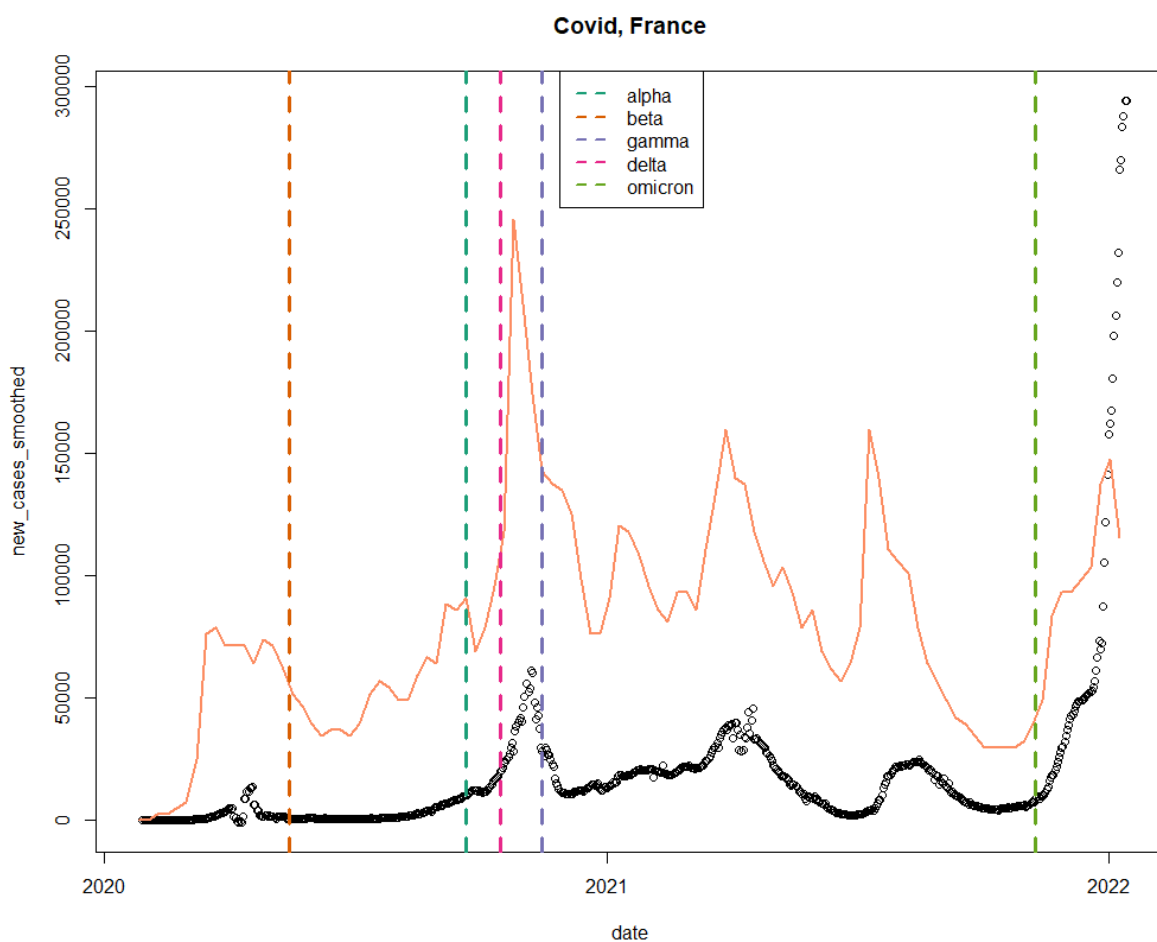


Рисунок 17 Графики популярности запросов по теме COVID-19 и подтвержденных случаев заболевания во Франции

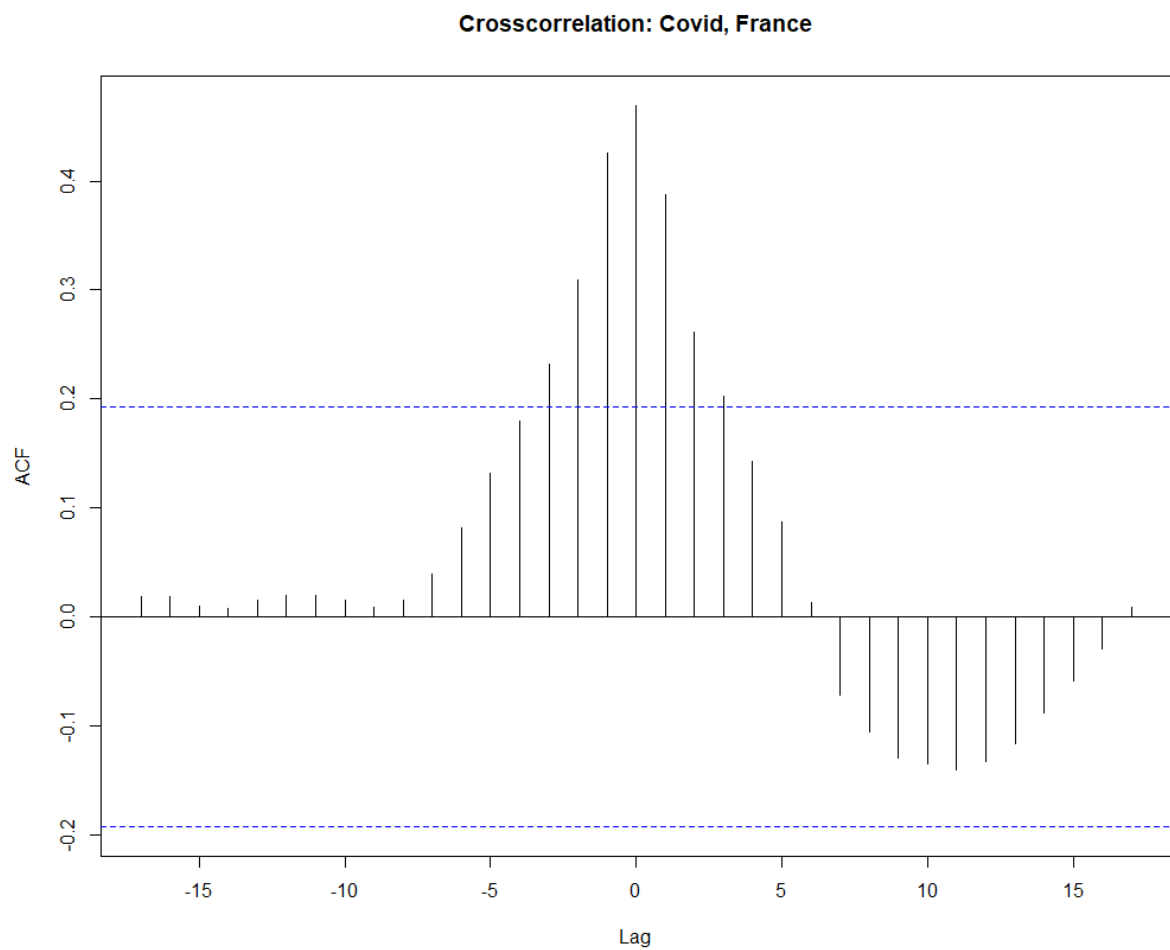


Рисунок 18 График кросскорреляции запросов по теме COVID-19 и заболеваемостью во Франции

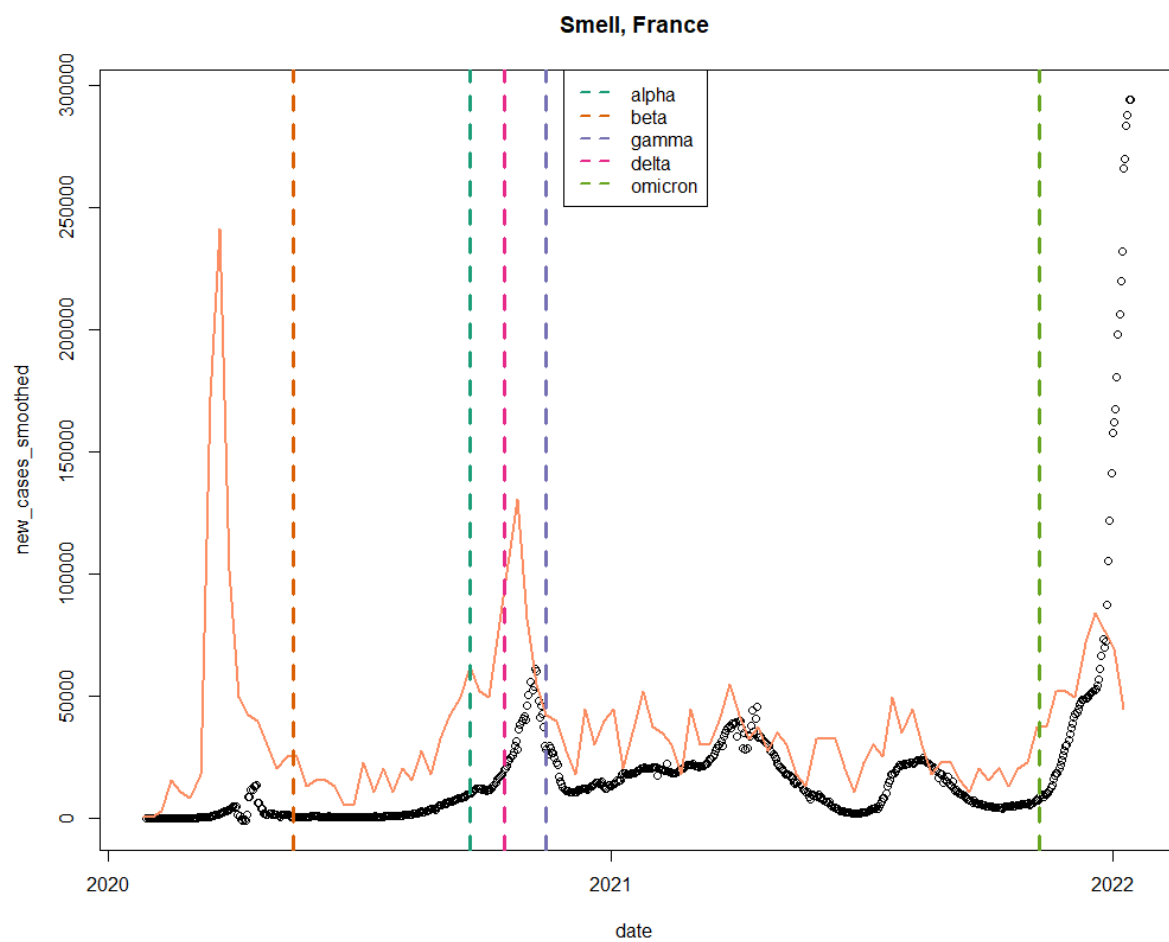


Рисунок 19 Графики популярности запросов по теме Обоняние и подтвержденных случаев заболевания во Франции

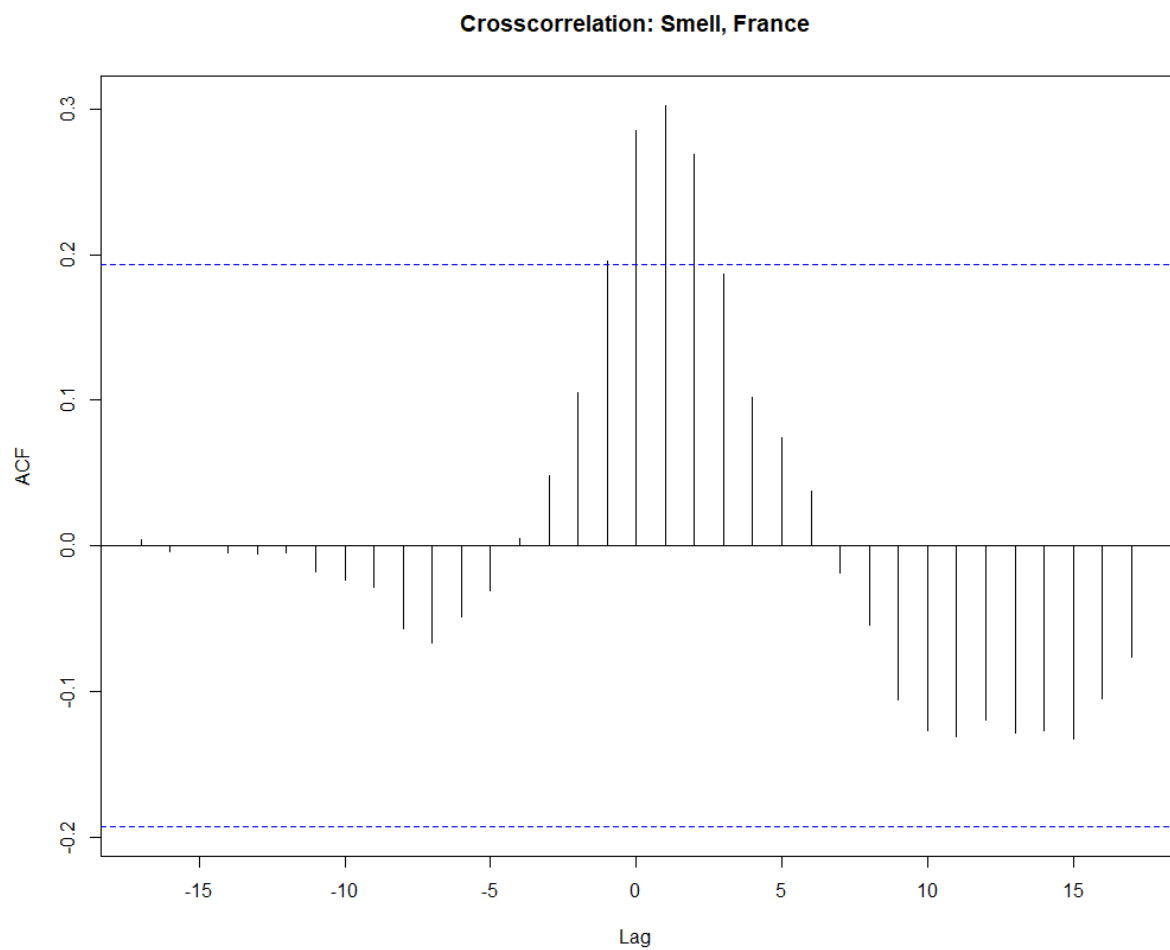


Рисунок 20 График кросскорреляции запросов по теме Обоняние и заболеваемостью во Франции

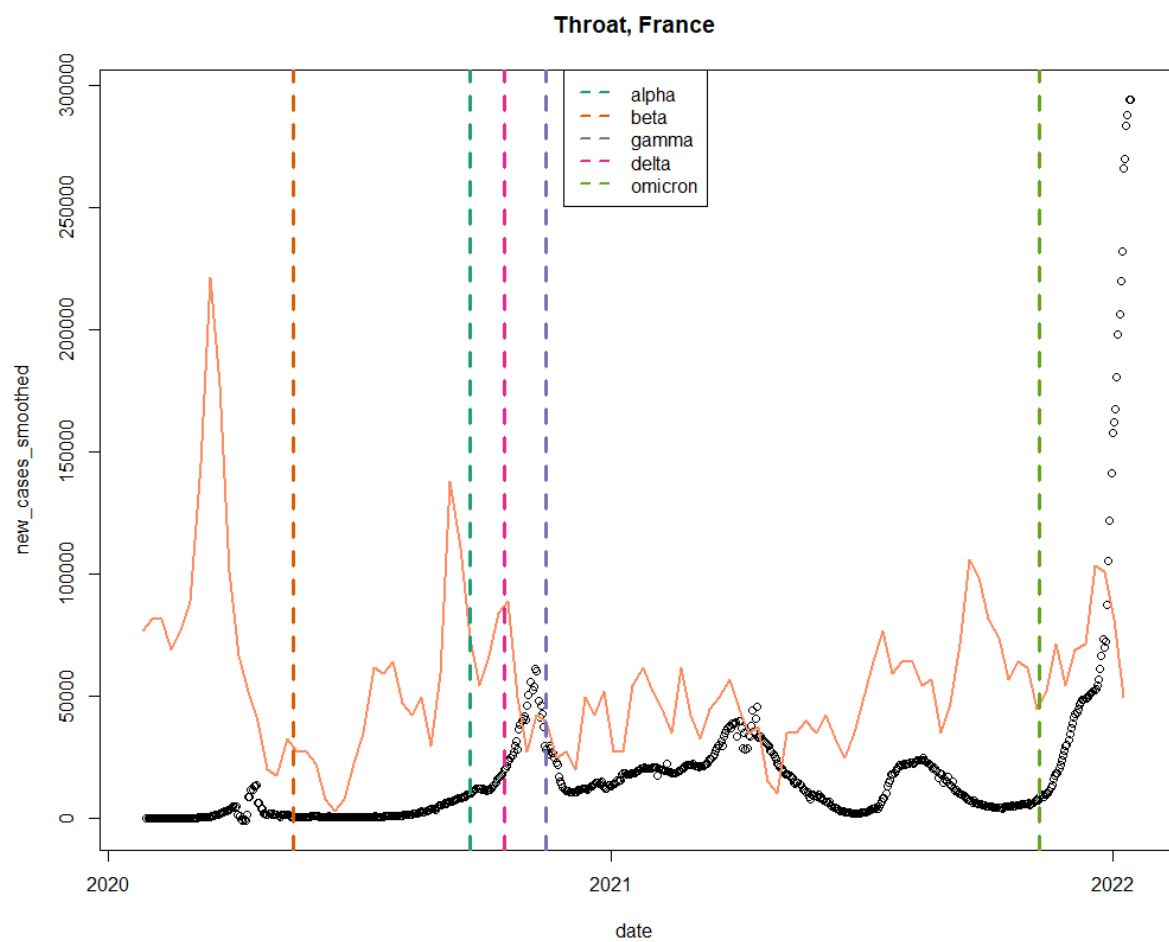


Рисунок 21 Графики популярности запросов по теме Больное горло и подтвержденных случаев заболевания во Франции

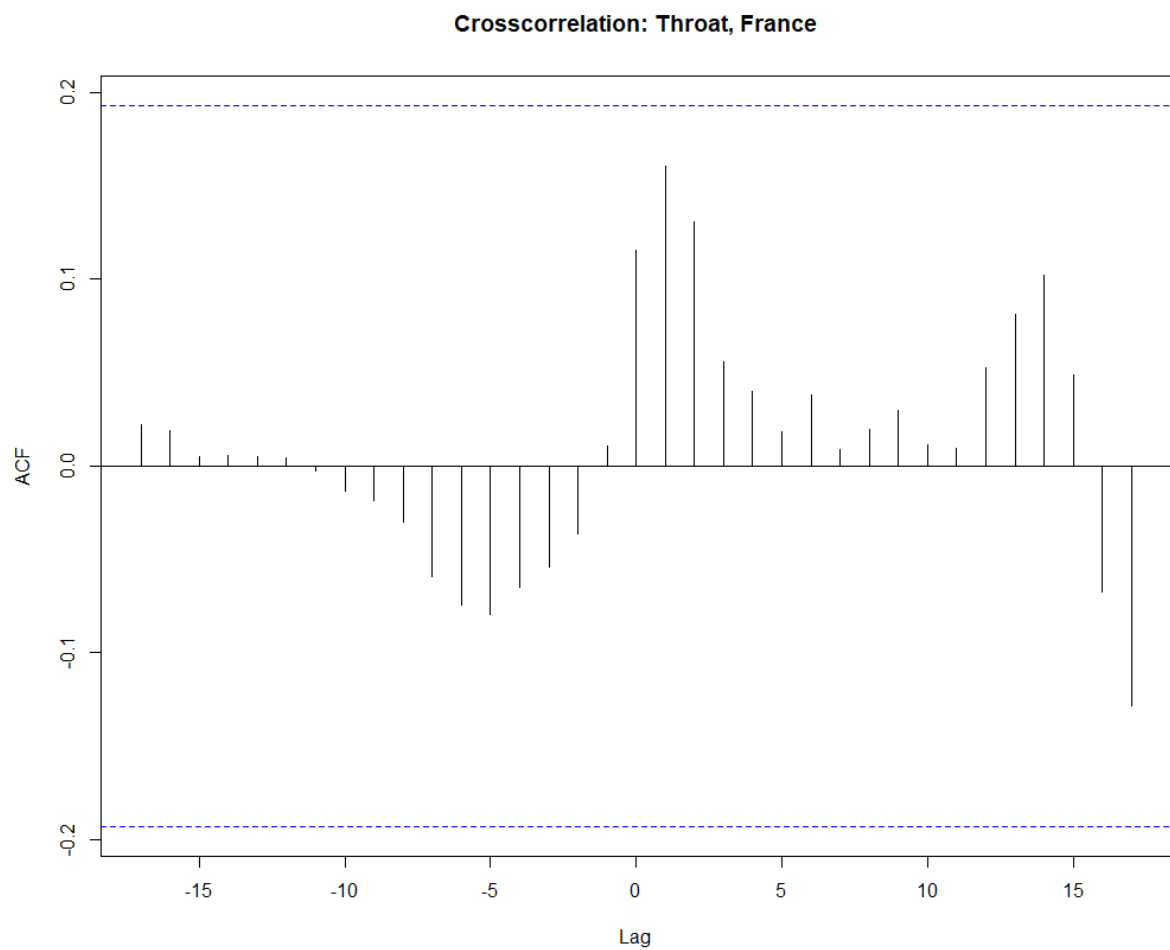


Рисунок 22 График кросскорреляции запросов по теме Боль в горле и заболеваемостью
во Франции

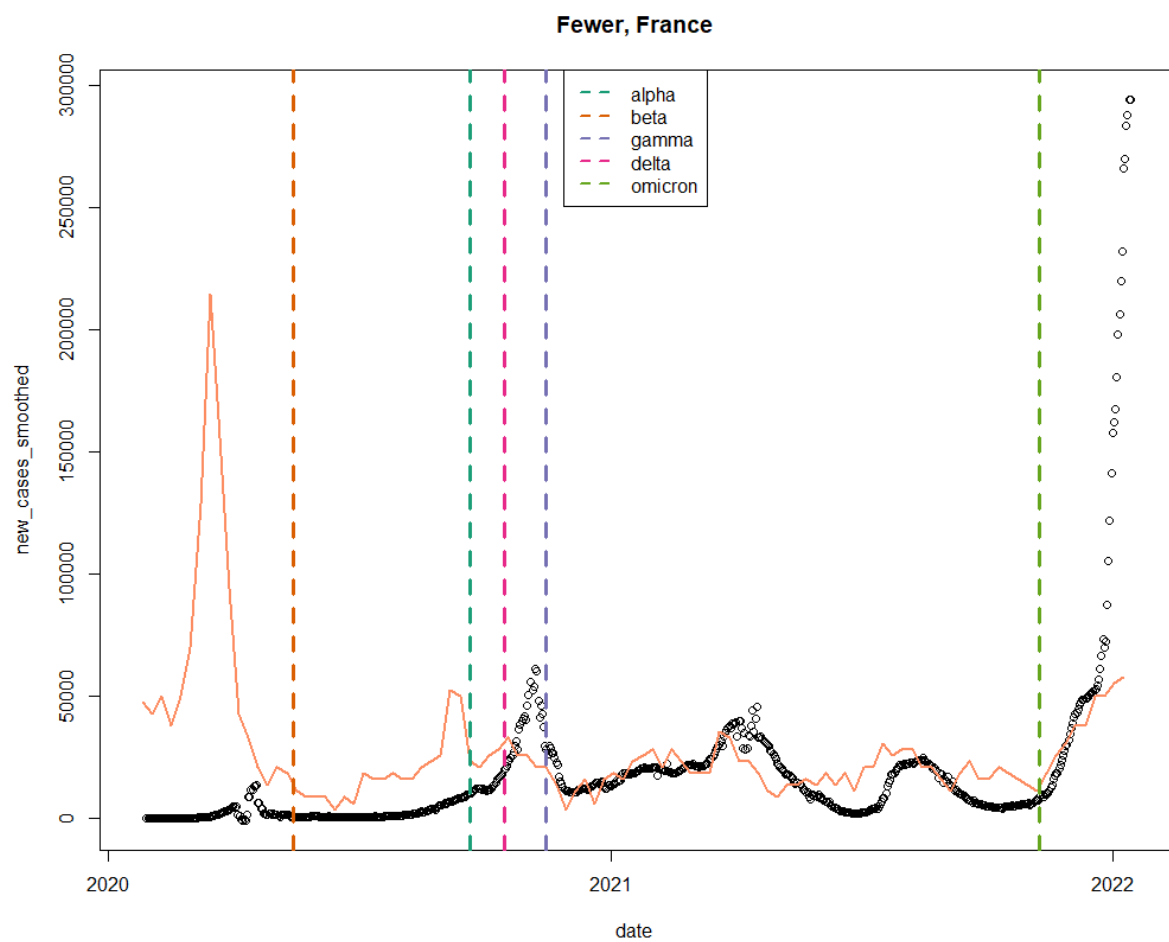


Рисунок 23 Графики популярности запросов по теме Лихорадка и подтвержденных случаев заболевания во Франции

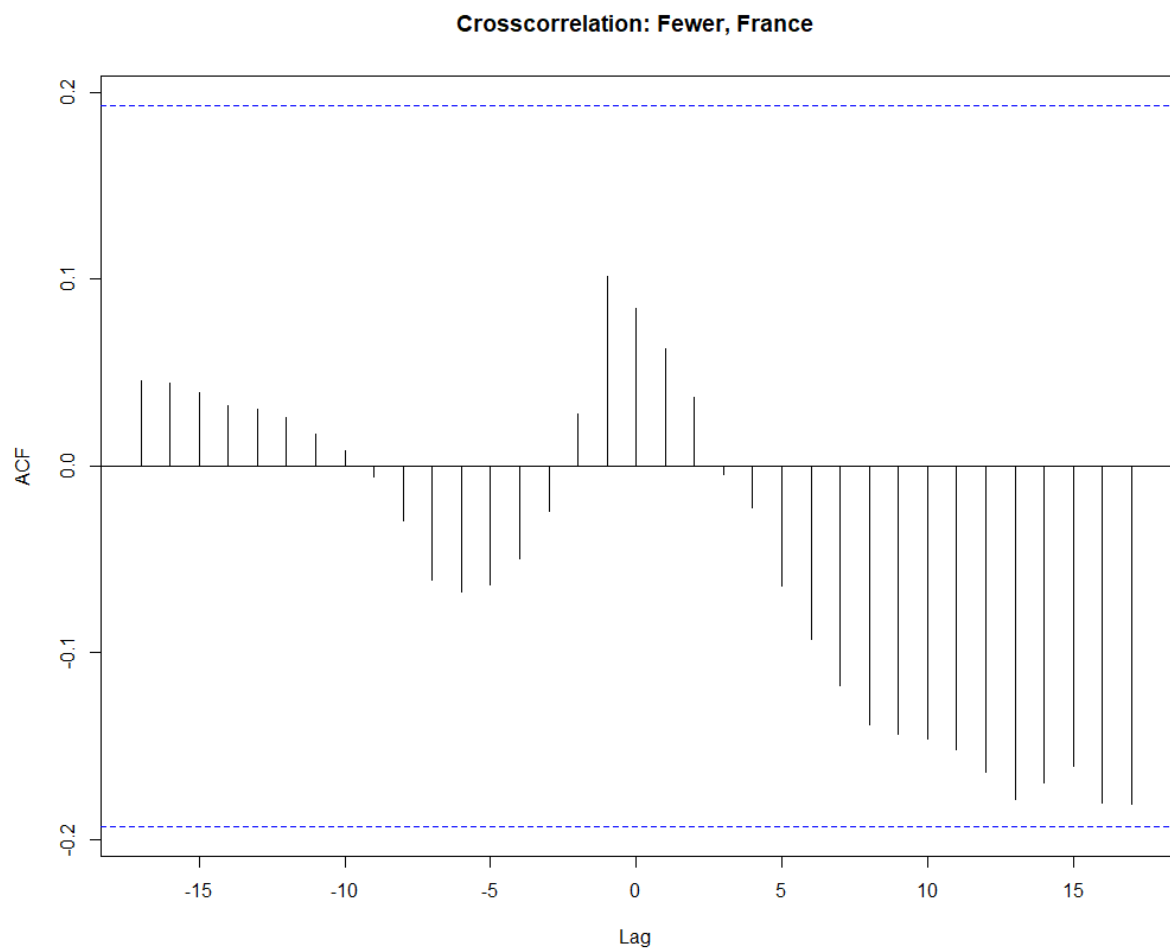


Рисунок 24 График кросскорреляции запросов по теме Лихорадка и заболеваемостью
во Франции

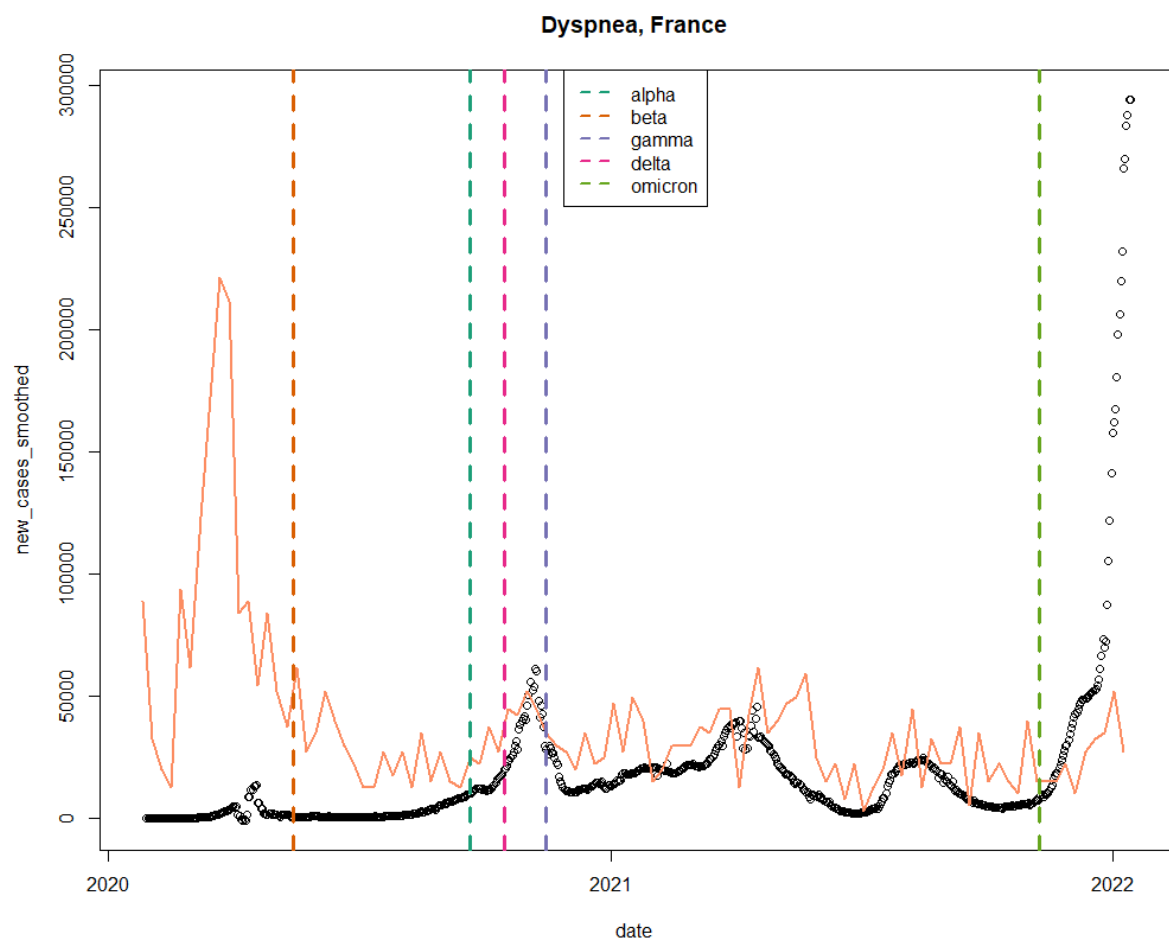


Рисунок 25 Графики популярности запросов по теме Одышка и подтвержденных случаев заболевания во Франции

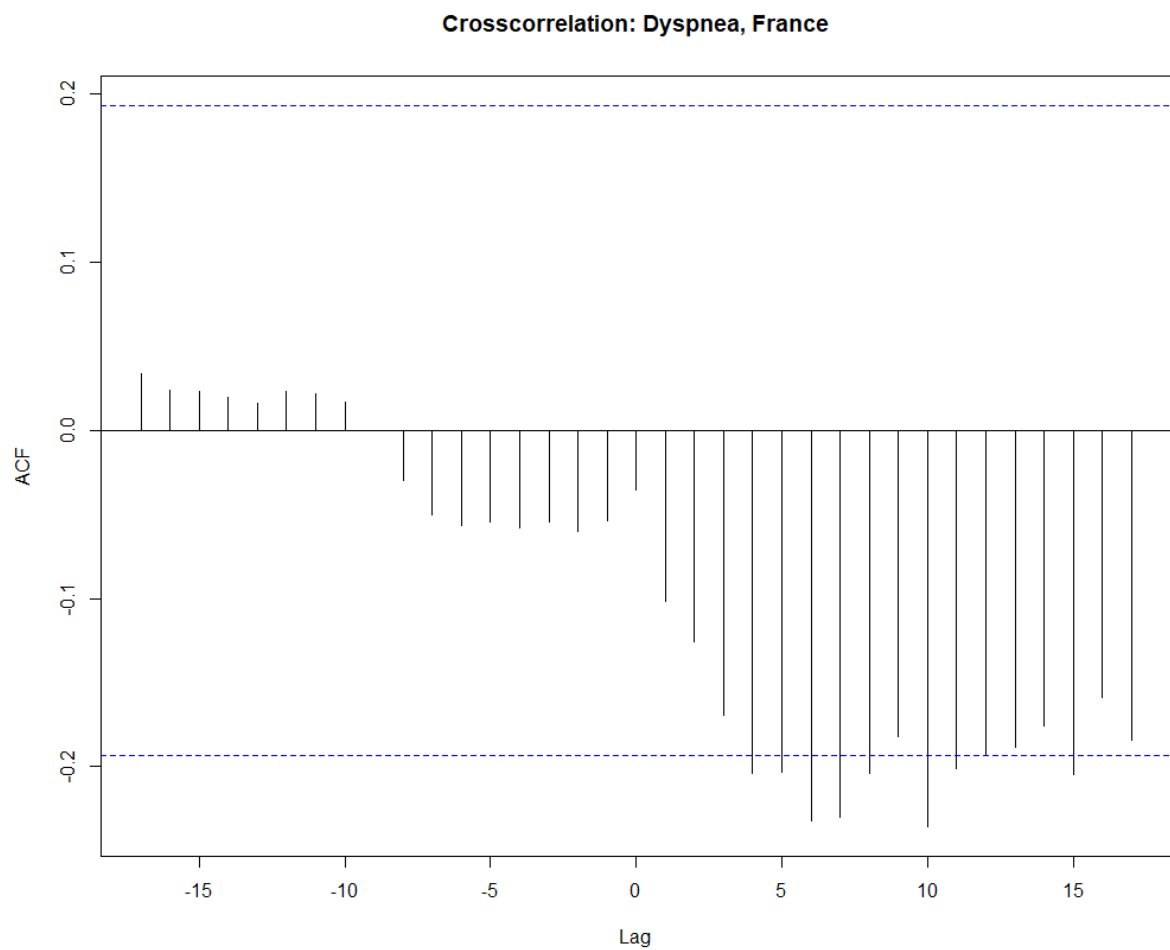


Рисунок 26 График кросскорреляции запросов по теме Одышка и заболеваемостью во Франции

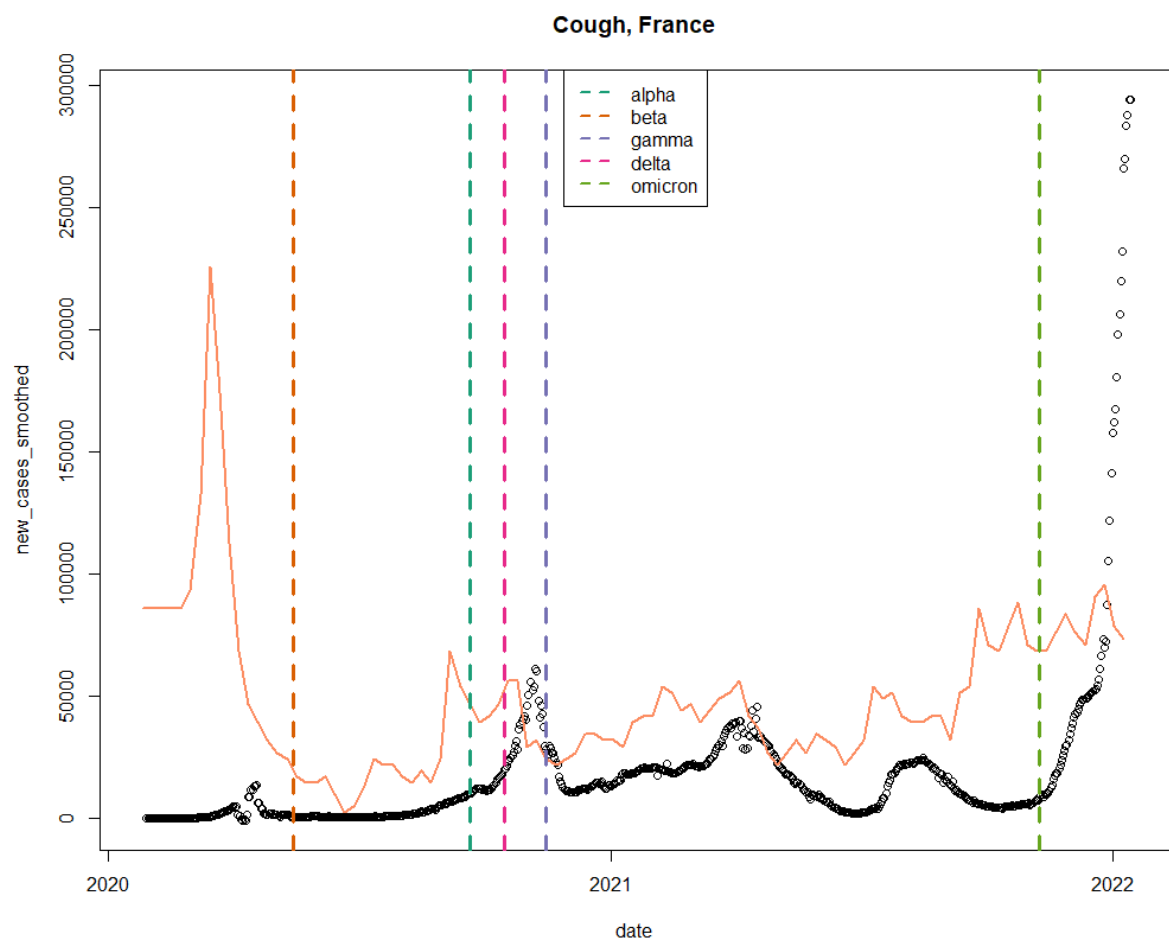


Рисунок 27 Графики популярности запросов по теме Кашель и подтвержденных случаев заболевания во Франции

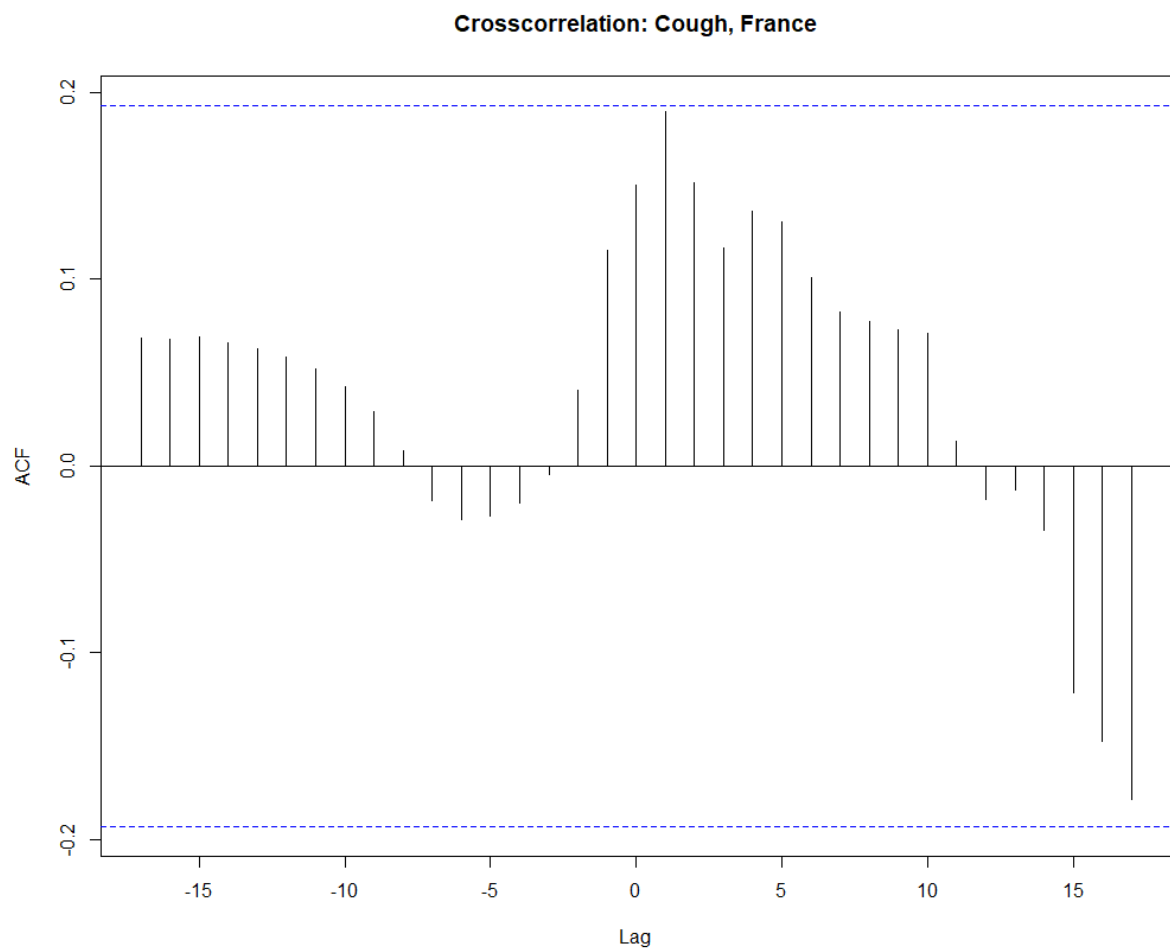


Рисунок 28 График кросскорреляции запросов по теме Кашель и заболеваемостью во Франции

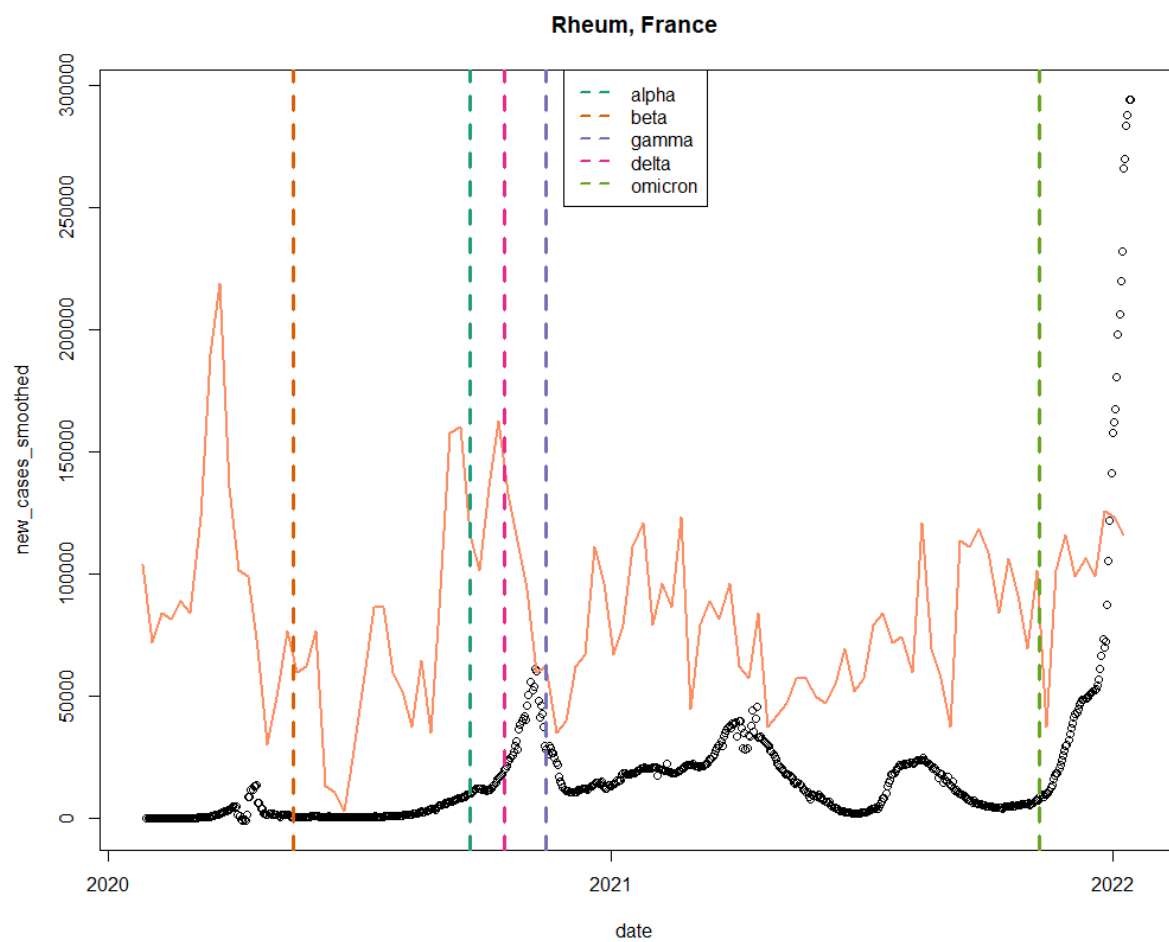


Рисунок 29 Графики популярности запросов по теме Насморк и подтвержденных случаев заболевания во Франции

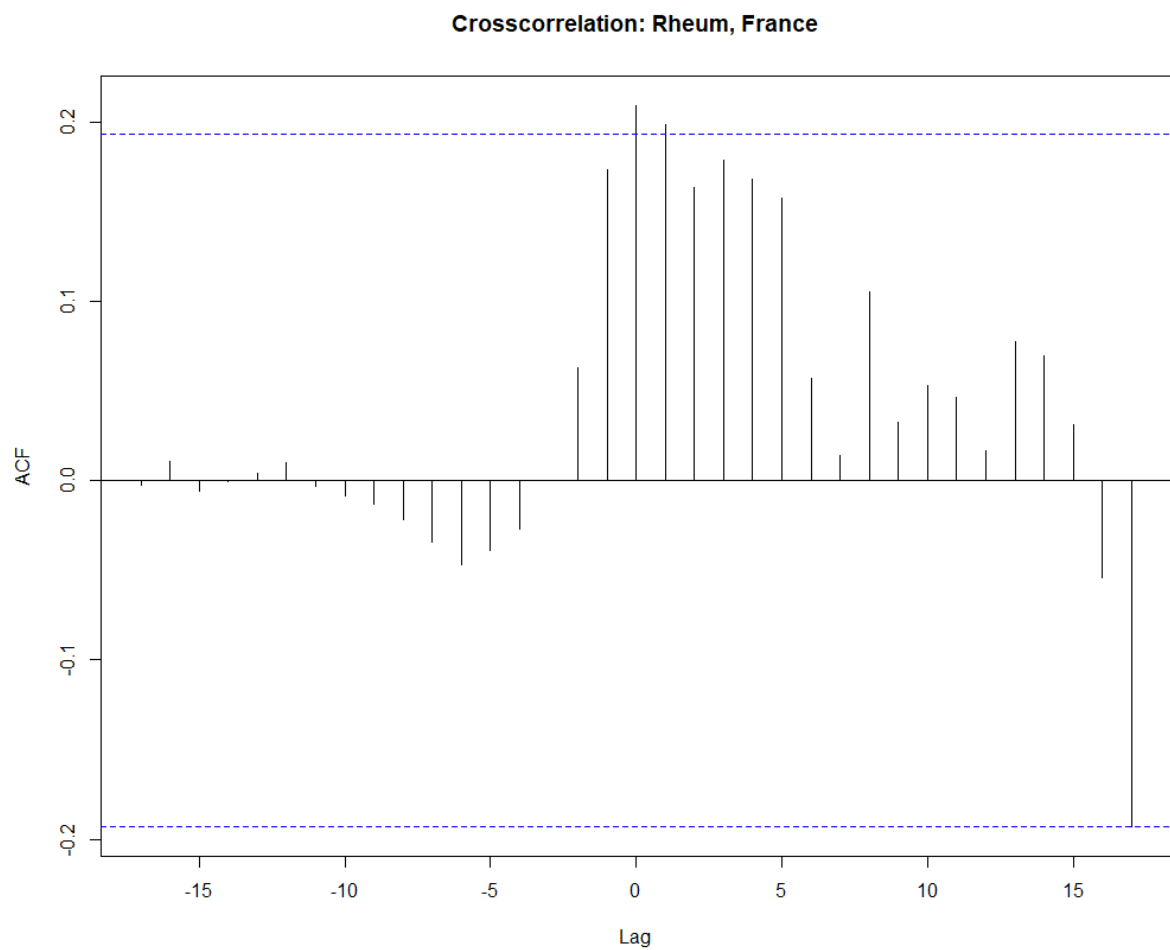


Рисунок 30 График кросскорреляции запросов по теме Насморк и заболеваемостью во Франции

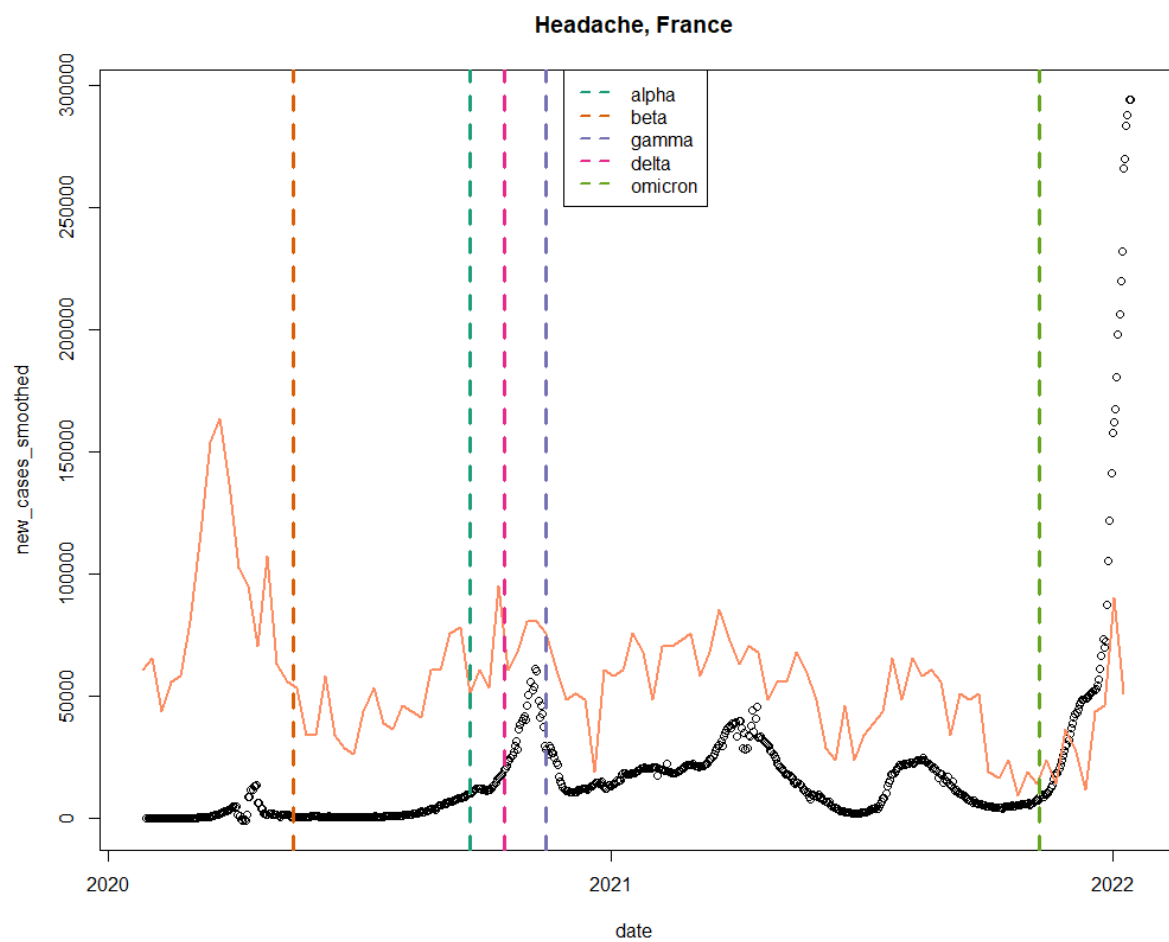


Рисунок 31 Графики популярности запросов по теме Обоняние и подтвержденных случаев заболевания во Франции

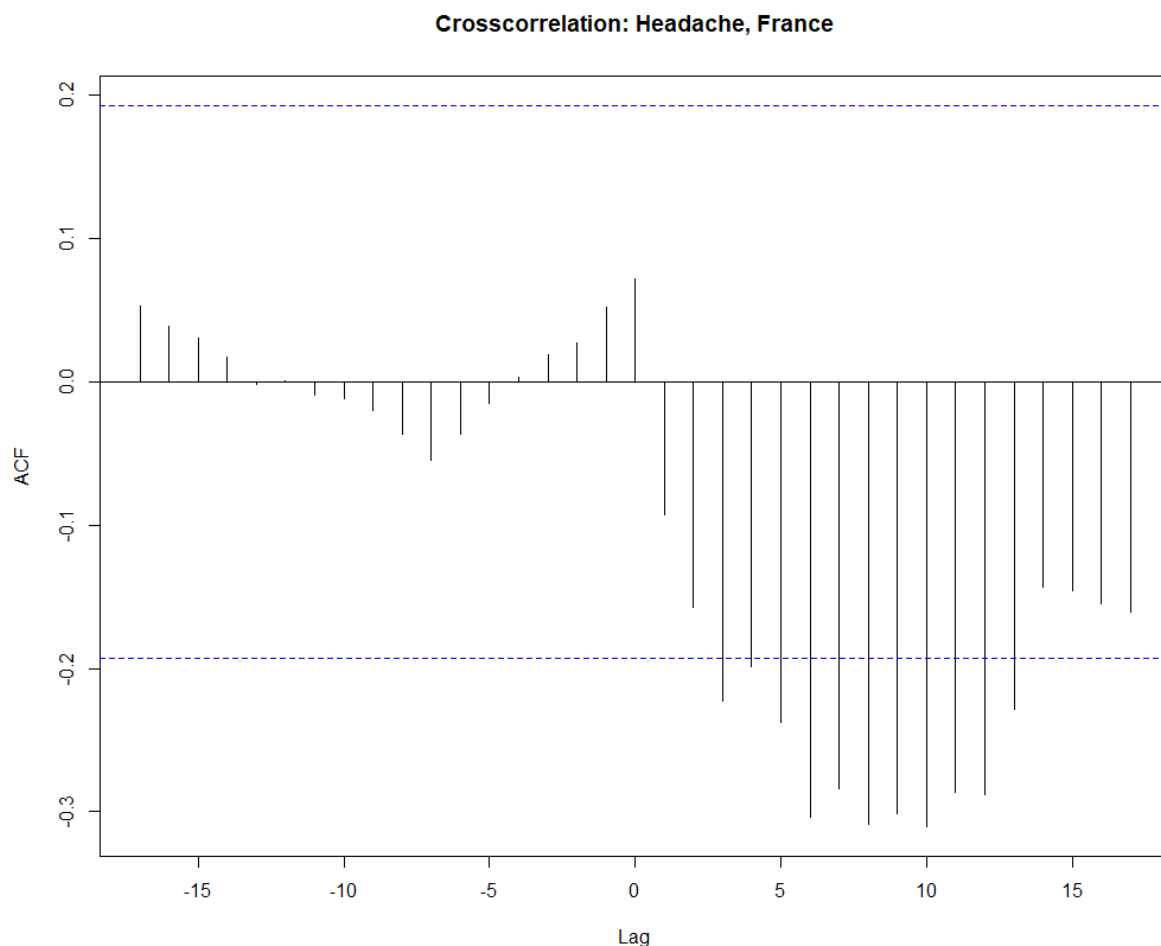


Рисунок 32 График кросскорреляции запросов по теме Головная боль и заболеваемостью во Франции

Анализ по стране

На графиках мы видим похожий сдвиг на второй волне, но пик кросскорреляционной функции смещен вправо намного меньше.

Почти на всех графиках мы видим аномалию в виде взрывного роста популярности рассматриваемых запросов, но при этом график заболеваемости почти не изменяется. Можно предположить, что в первую волну система диагностики заболевания не справилась с количеством заболевших.

США

Таблица 4 Корреляция заболеваемости и популярности запросов в США

WORDKEY	PEARSON	SPEARMAN
COVID	-0.06182	0.119239
SMELL	0.492288	0.597831
THROAT	0.211912	0.095149
FEWER	0.183164	0.101951
DYSPNEA	-0.1734	-0.08512
COUGH	0.168117	0.072499
RHEUM	0.518913	0.348671
HEADACHE	0.324957	0.245963

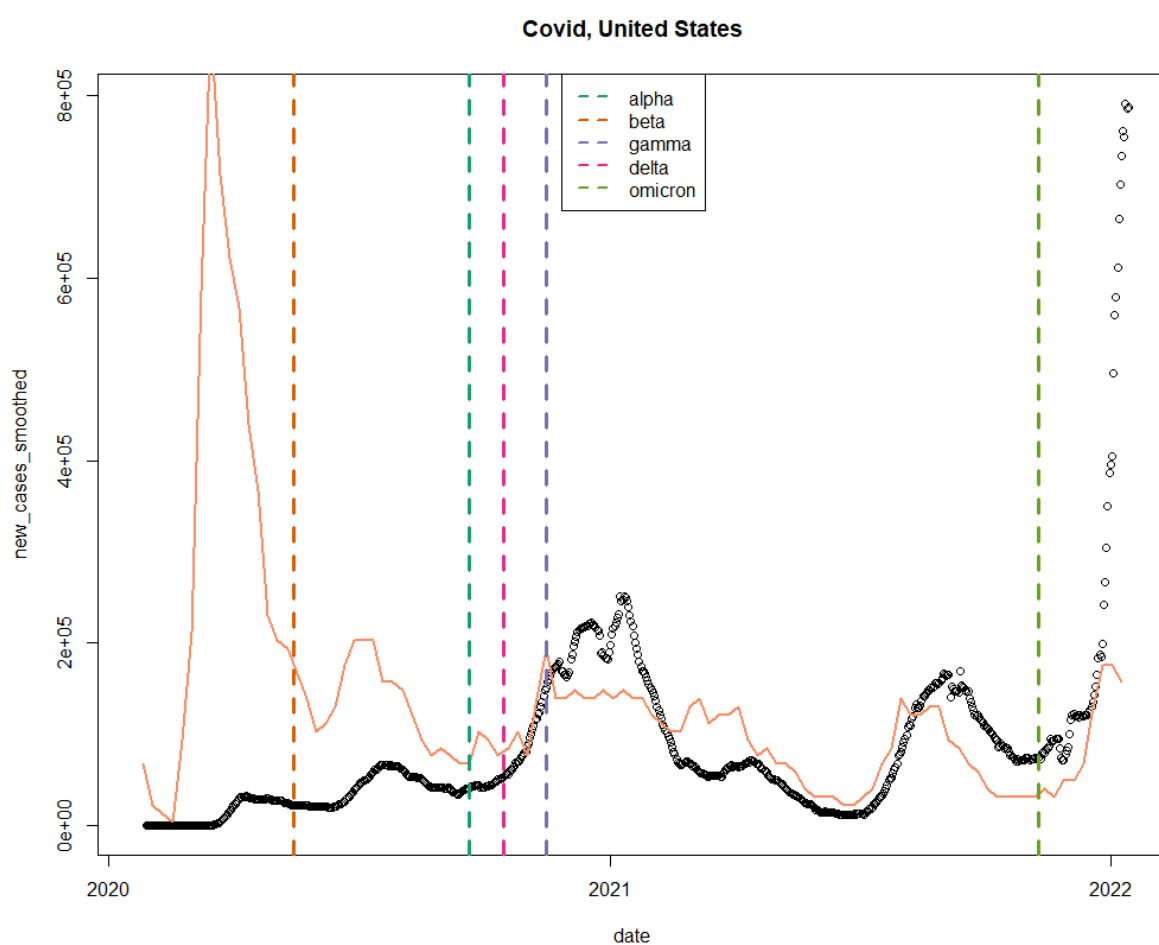


Рисунок 33 Графики популярности запросов по теме COVID-19 и подтвержденных случаев заболевания в США

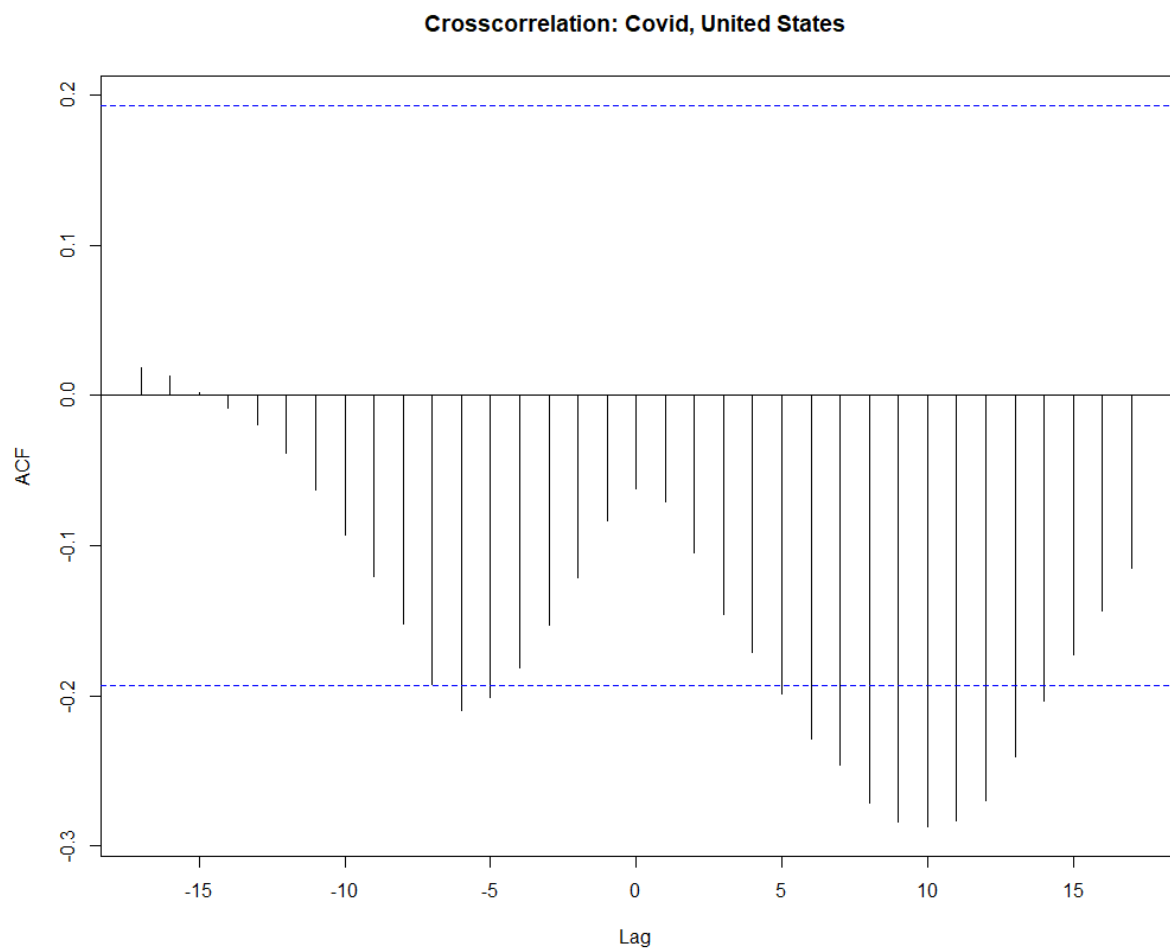


Рисунок 34 График кросскорреляции запросов по теме COVID-19 и заболеваемостью в США

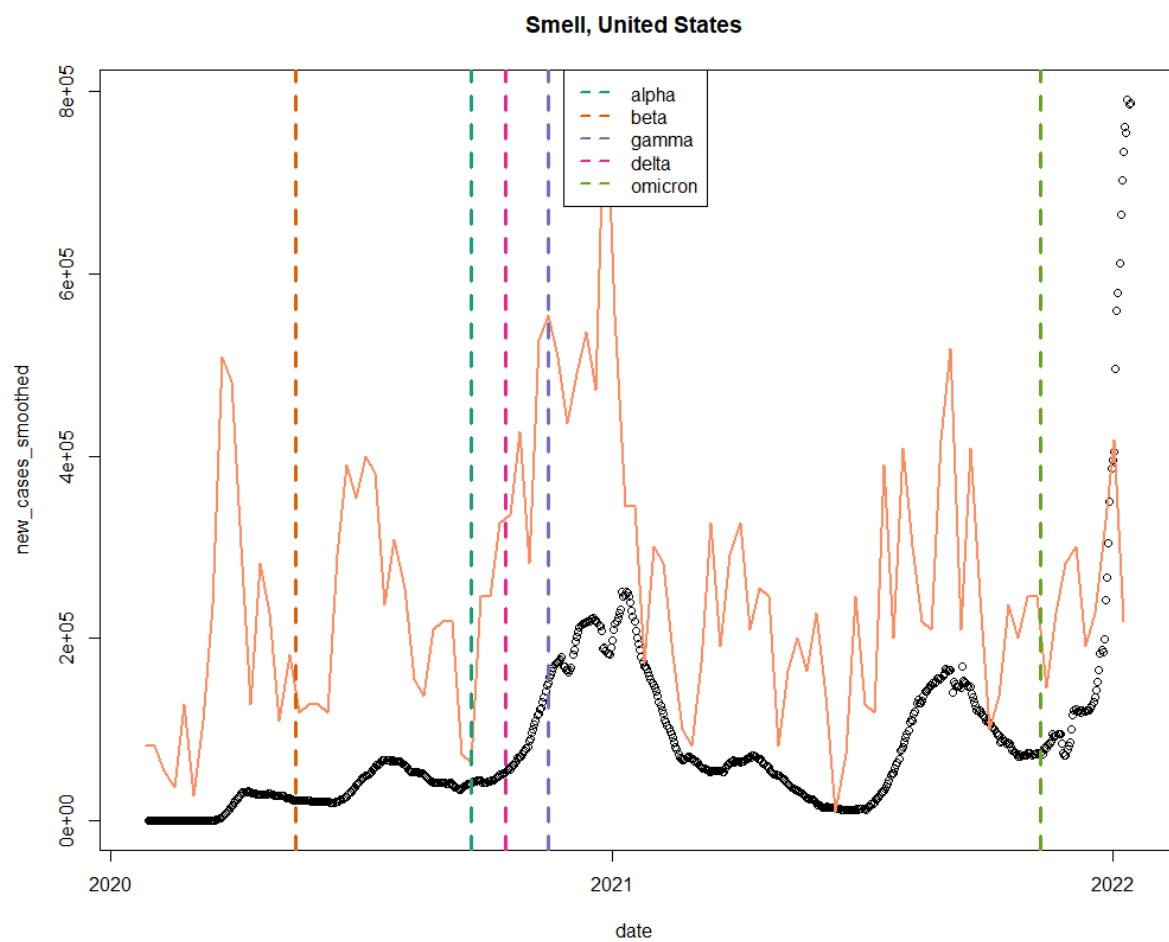


Рисунок 35 Графики популярности запросов по теме Обоняние и подтвержденных случаев заболевания в США

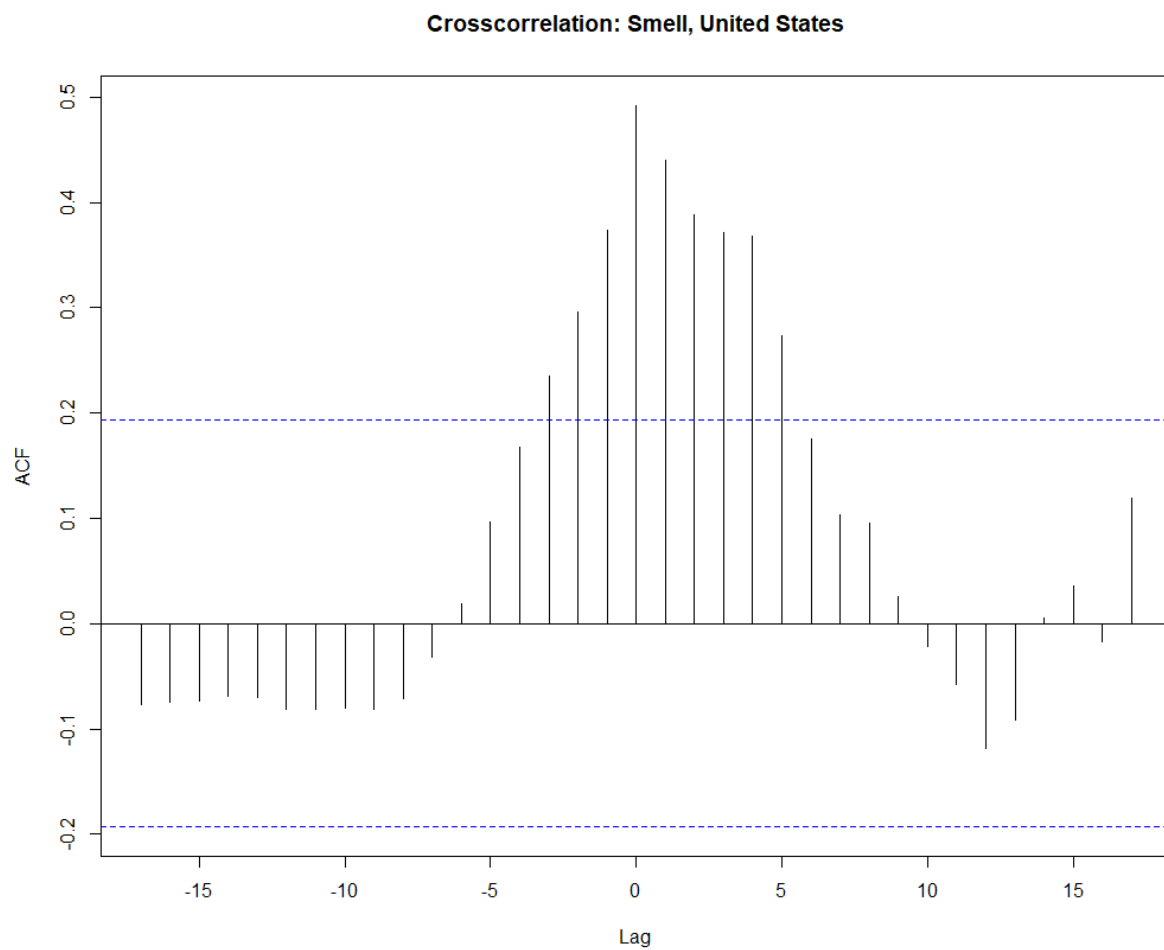


Рисунок 36 График кросскорреляции запросов по теме Обоняние и заболеваемостью в США

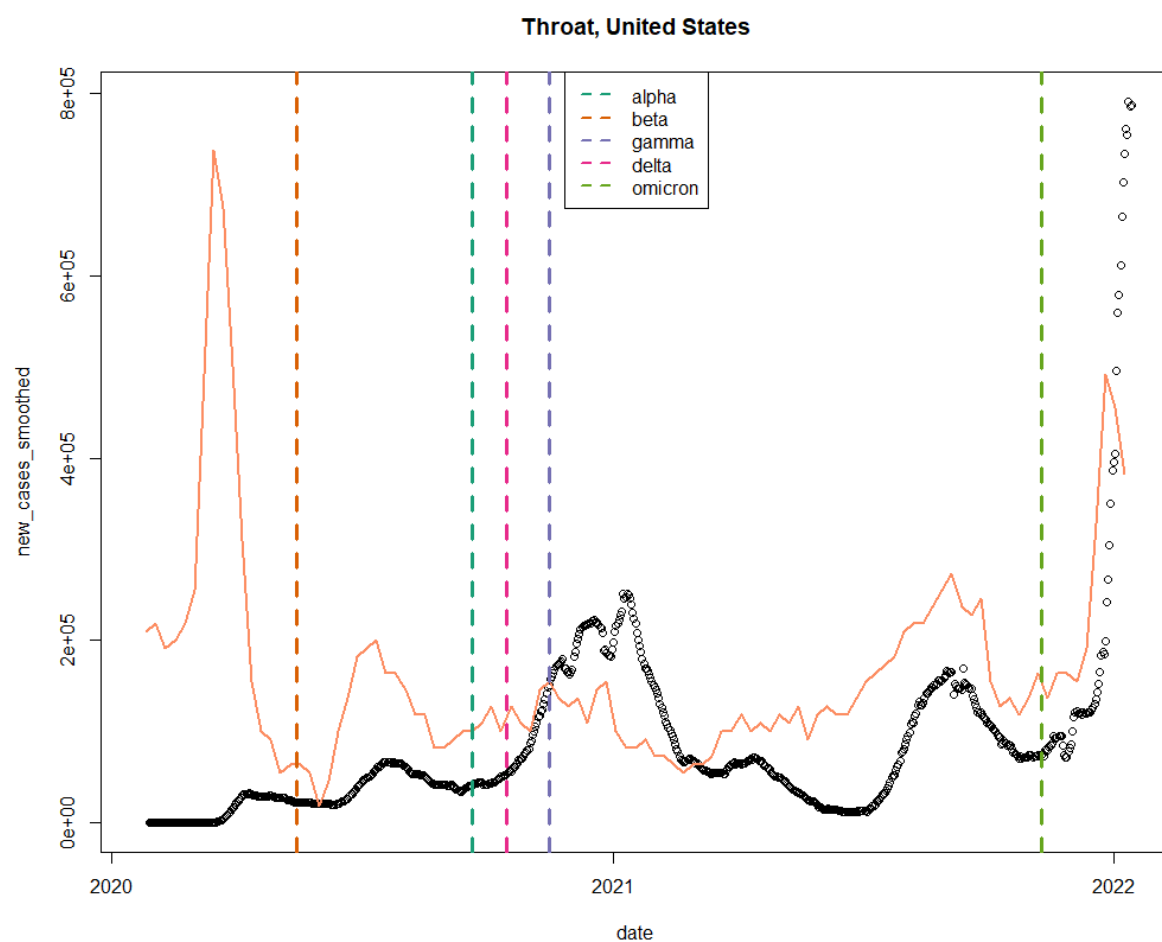


Рисунок 37 Графики популярности запросов по теме Больное горло и подтвержденных случаев заболевания в США

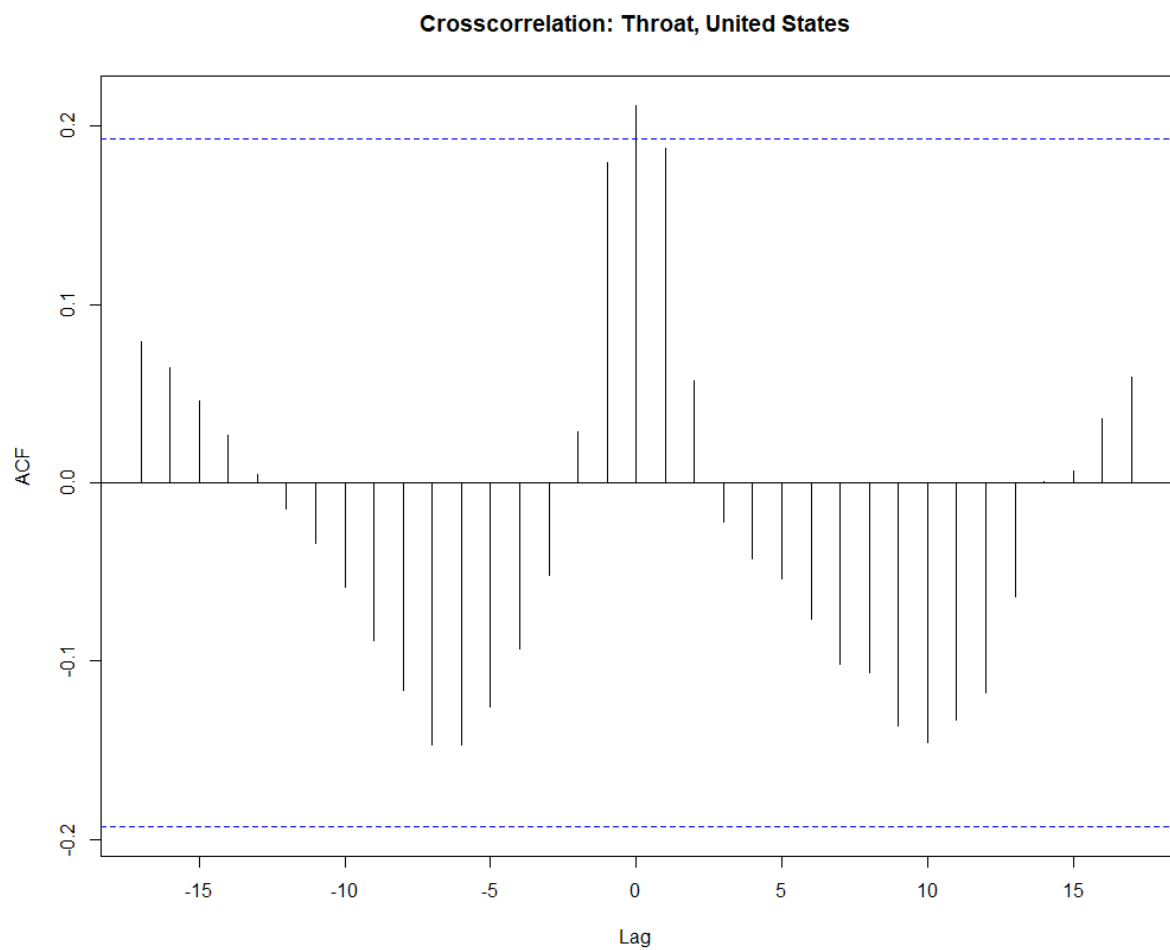


Рисунок 38 График кросскорреляции запросов по теме Боль в горле в США

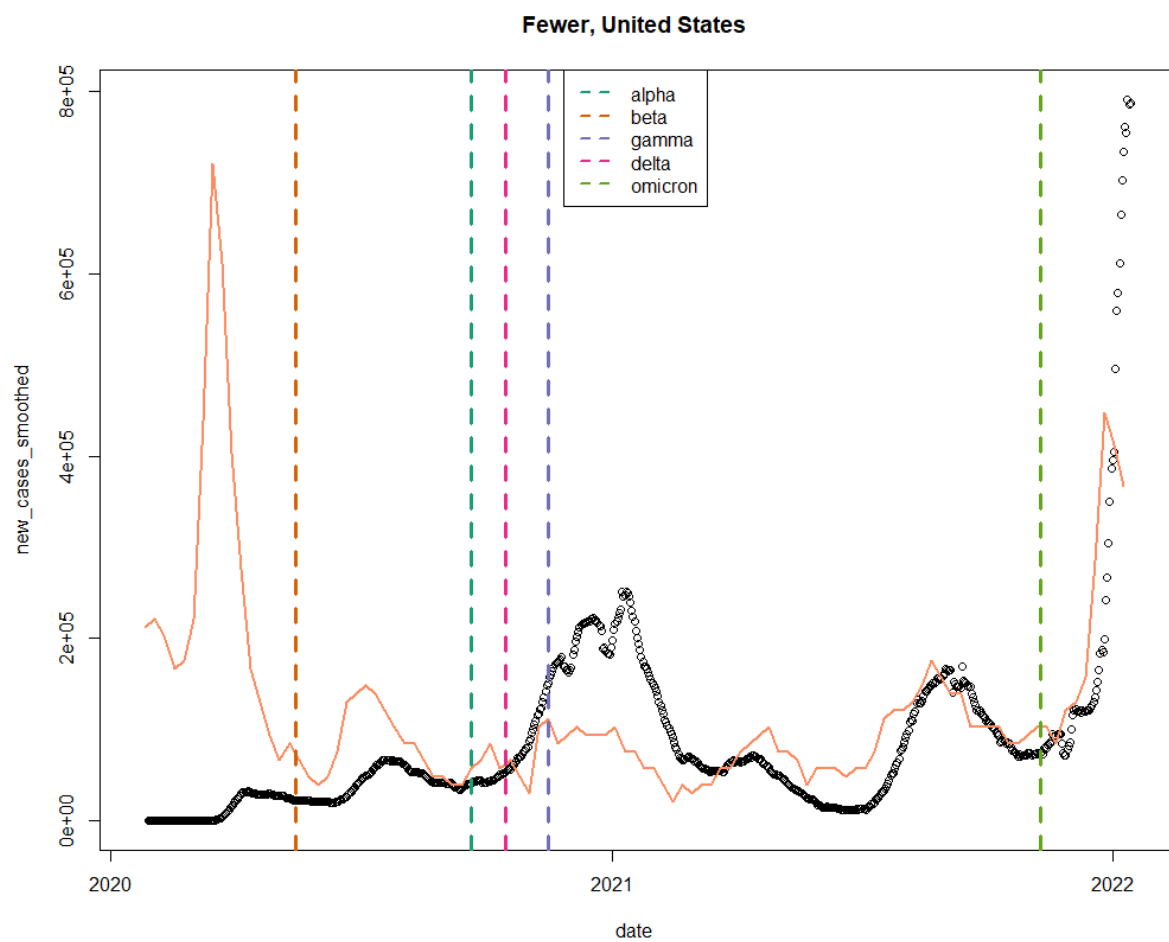


Рисунок 39 Графики популярности запросов по теме Лихорадка и подтвержденных случаев заболевания в США

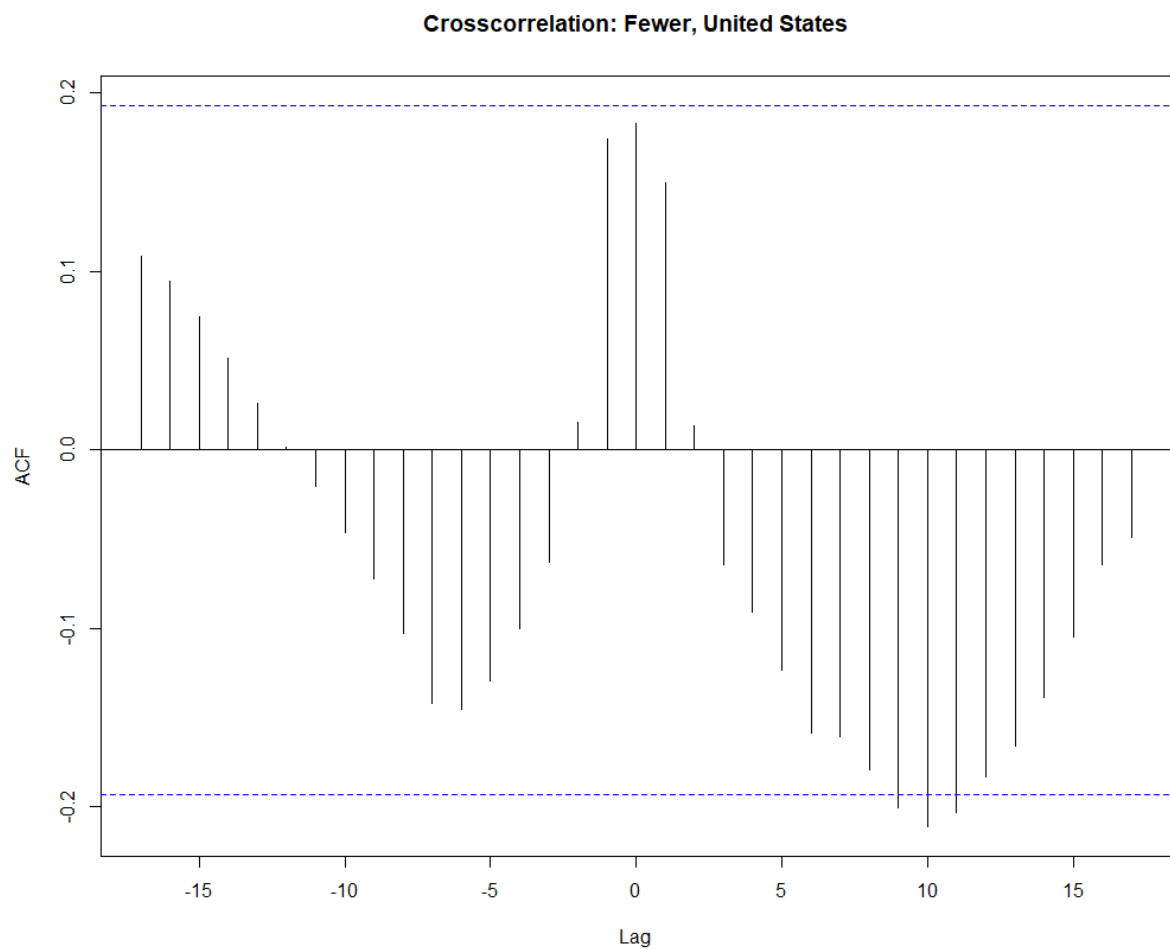


Рисунок 40 График кросскорреляции запросов по теме Лихорадка и заболеваемостью в США

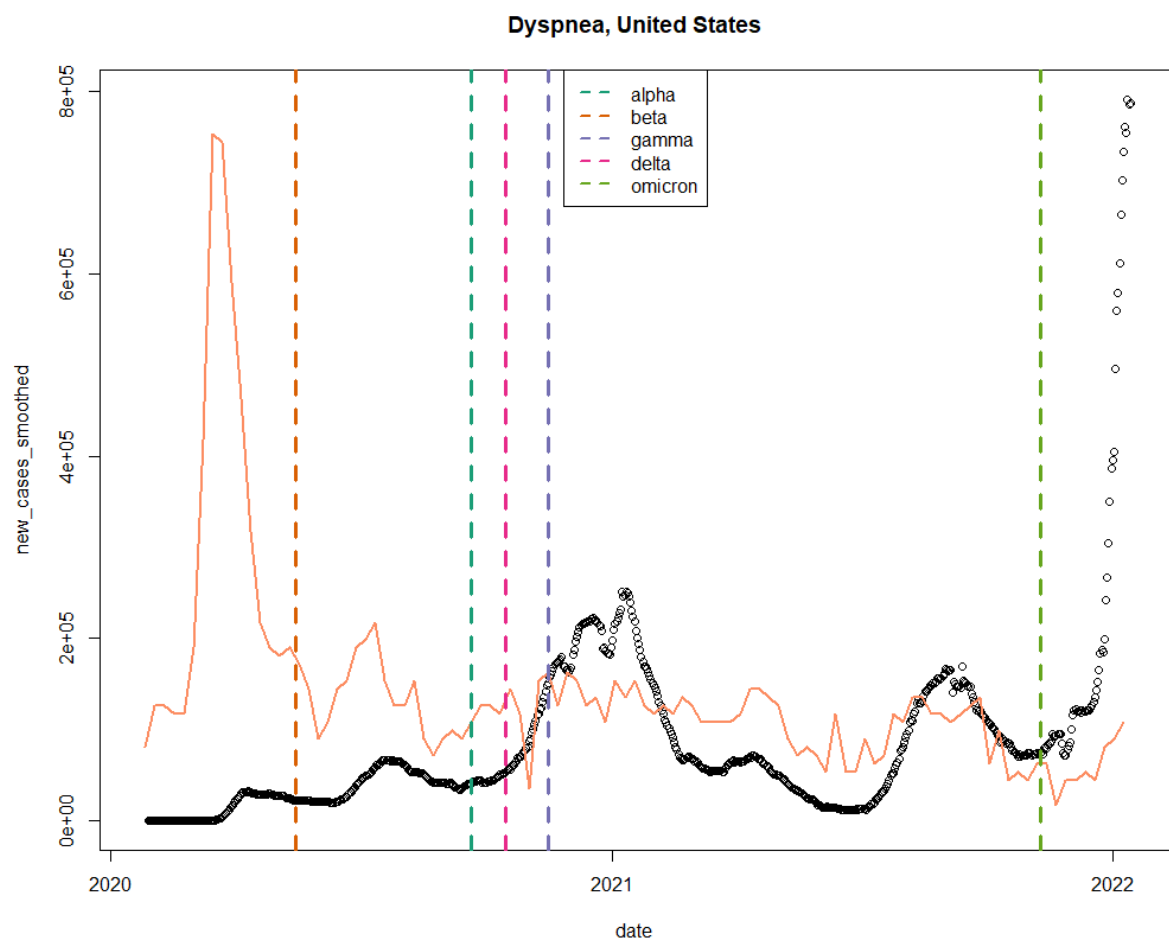


Рисунок 41 Графики популярности запросов по теме Одышка и подтвержденных случаев заболевания в США

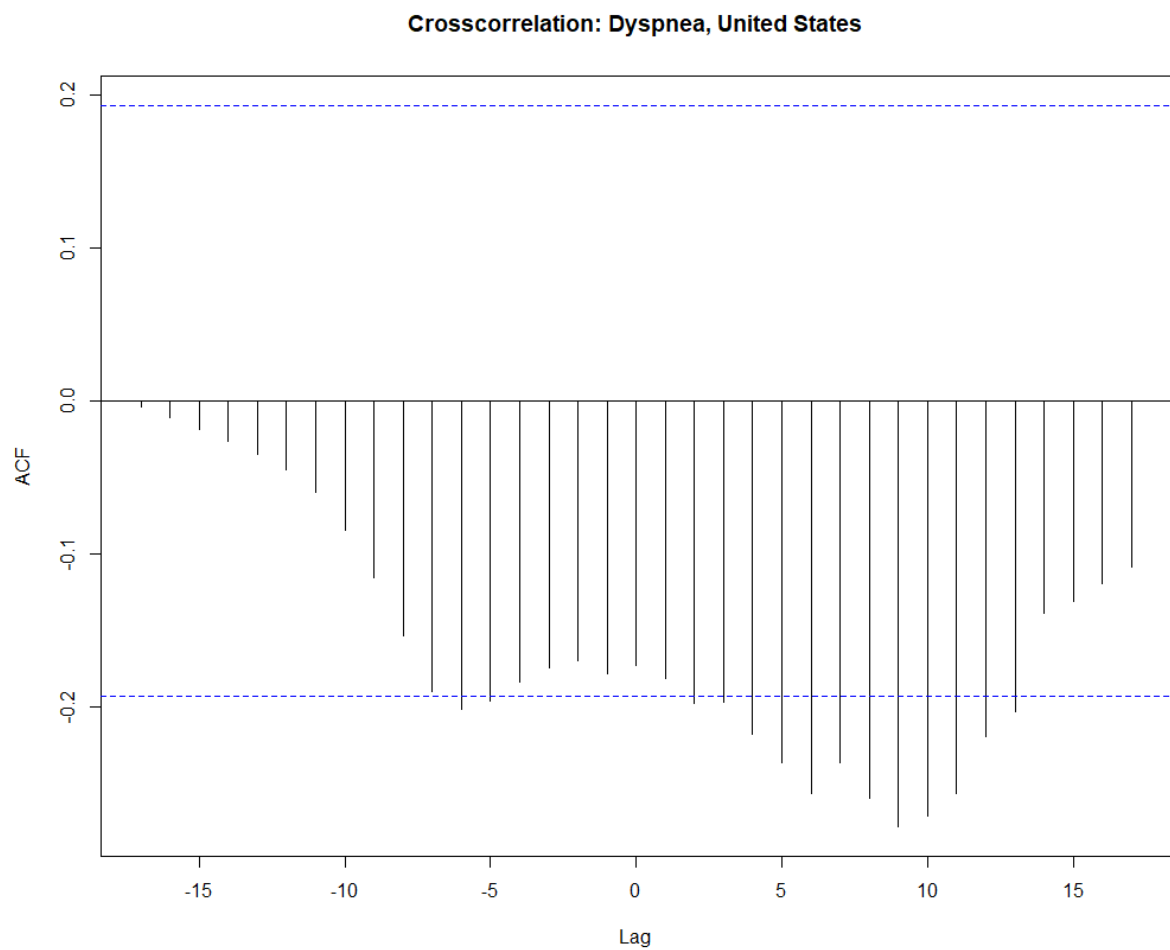


Рисунок 42 График кросскорреляции запросов по теме Одышка и заболеваемостью в США

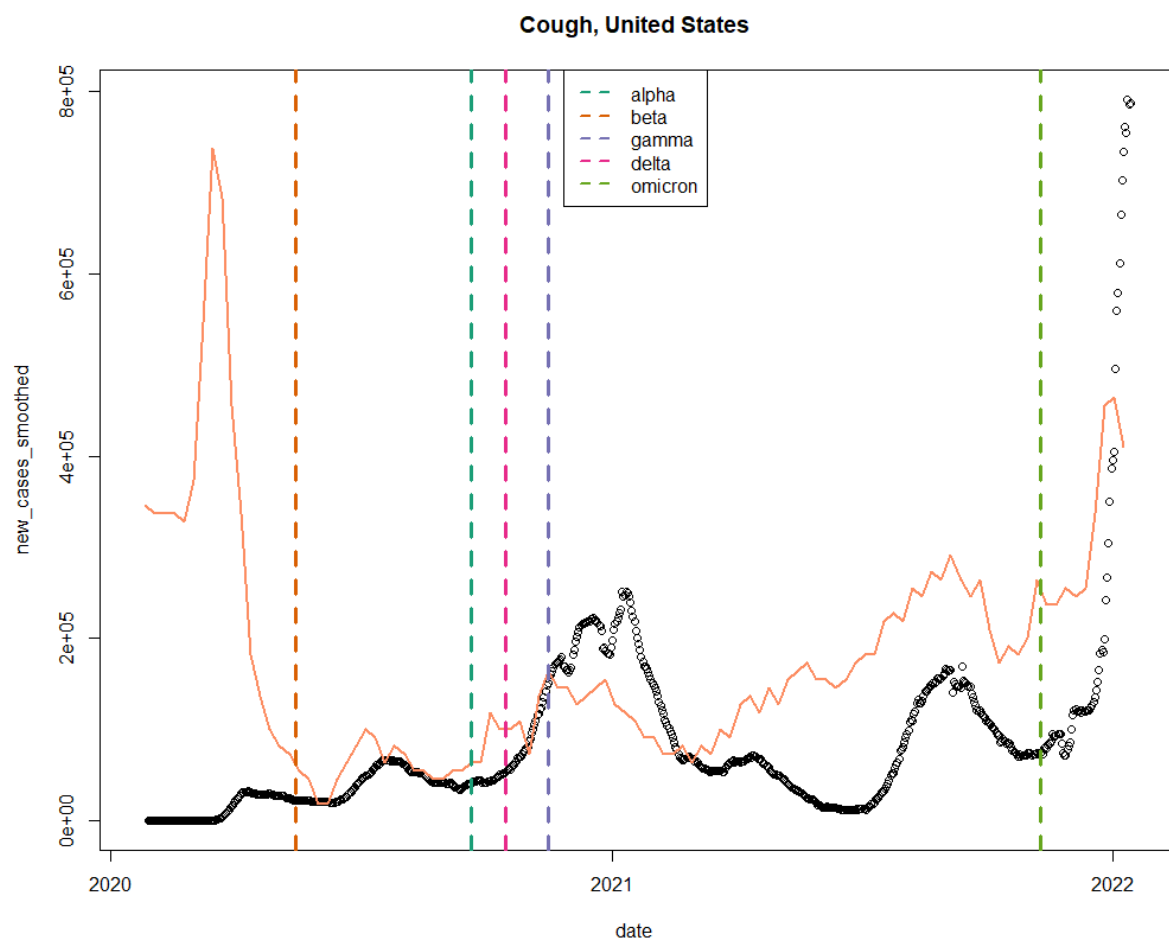


Рисунок 43 Графики популярности запросов по теме Кашель и подтвержденных случаев заболевания в США

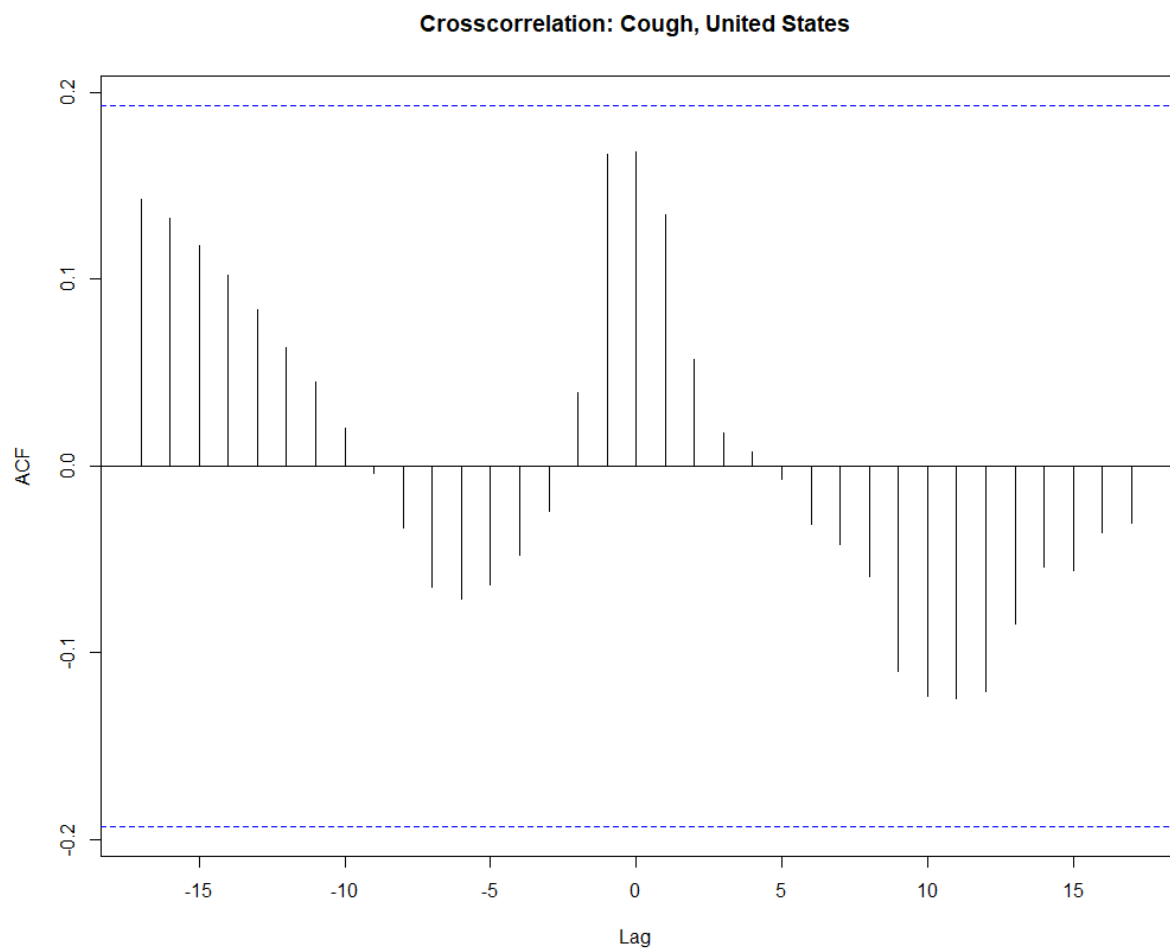


Рисунок 44 График кросскорреляции запросов по теме Кашель и заболеваемостью в США

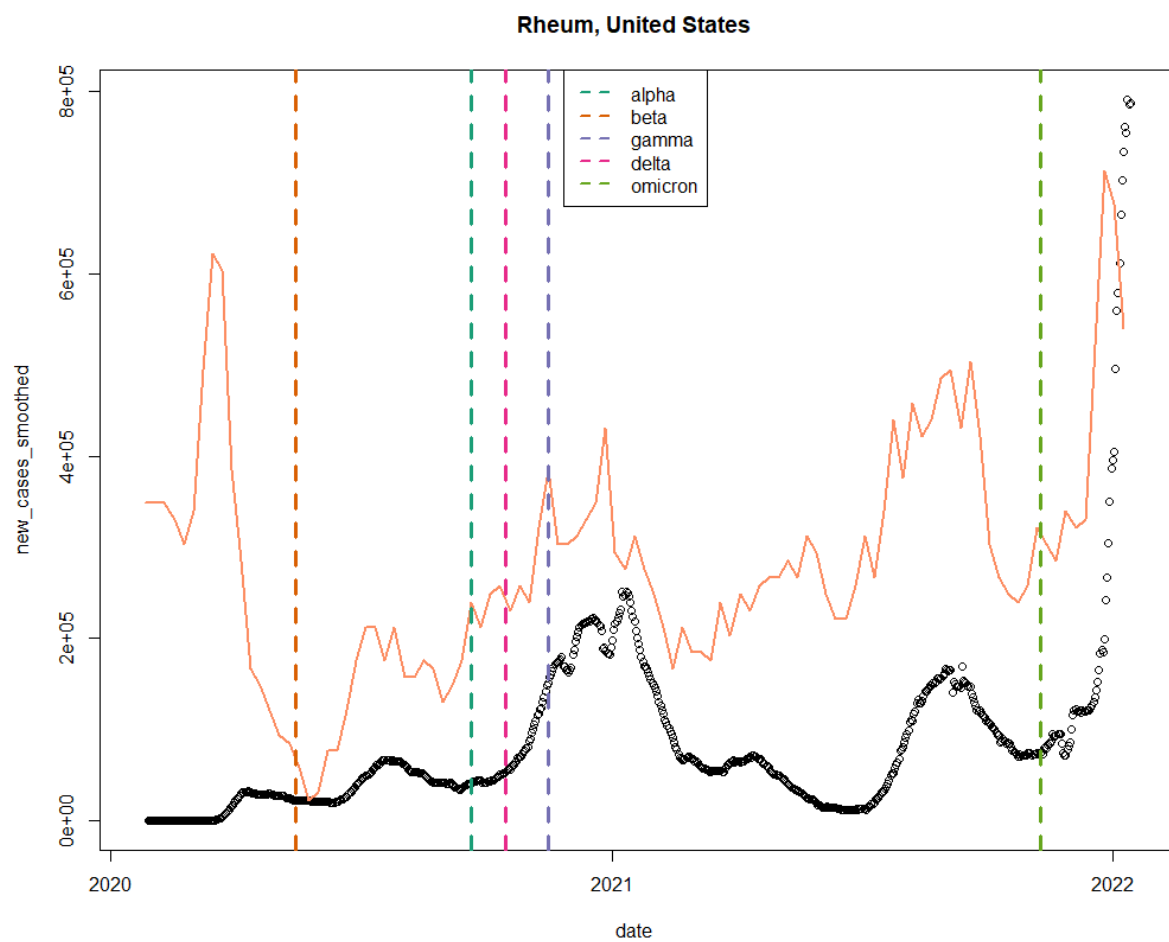


Рисунок 45 Графики популярности запросов по теме Насморк и подтвержденных случаев заболевания в США

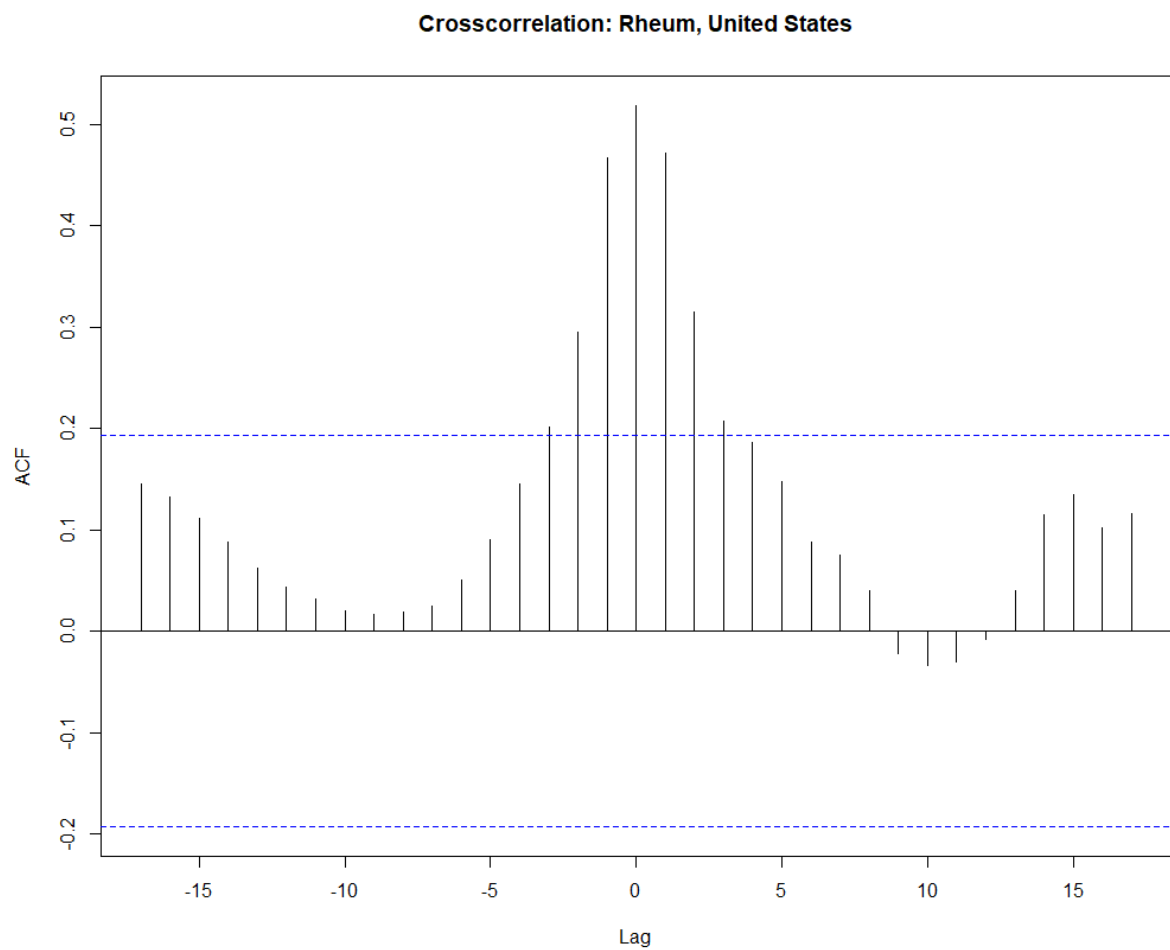


Рисунок 46 График кросскорреляции запросов по теме Насморк и заболеваемостью в США

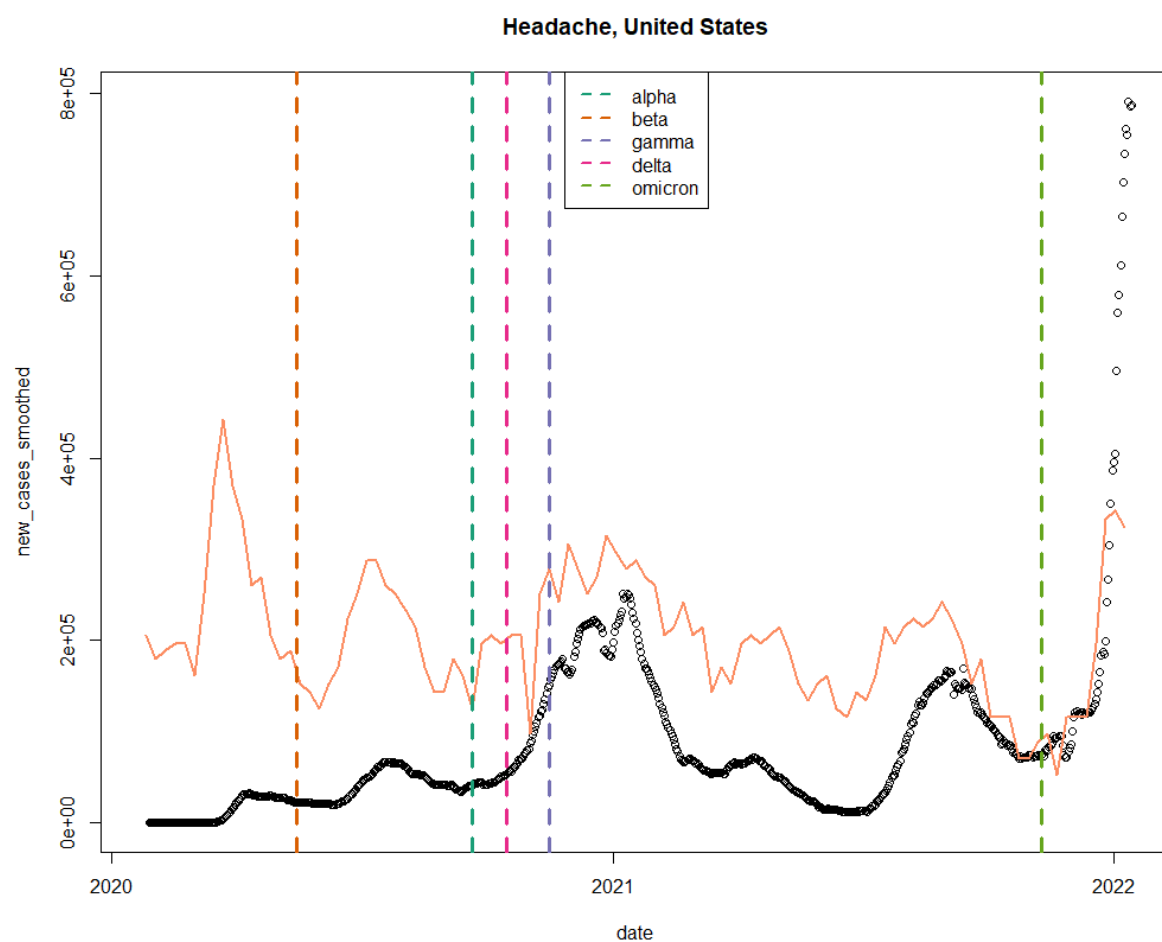


Рисунок 47 Графики популярности запросов по теме Головная боль и подтвержденных случаев заболевания в США

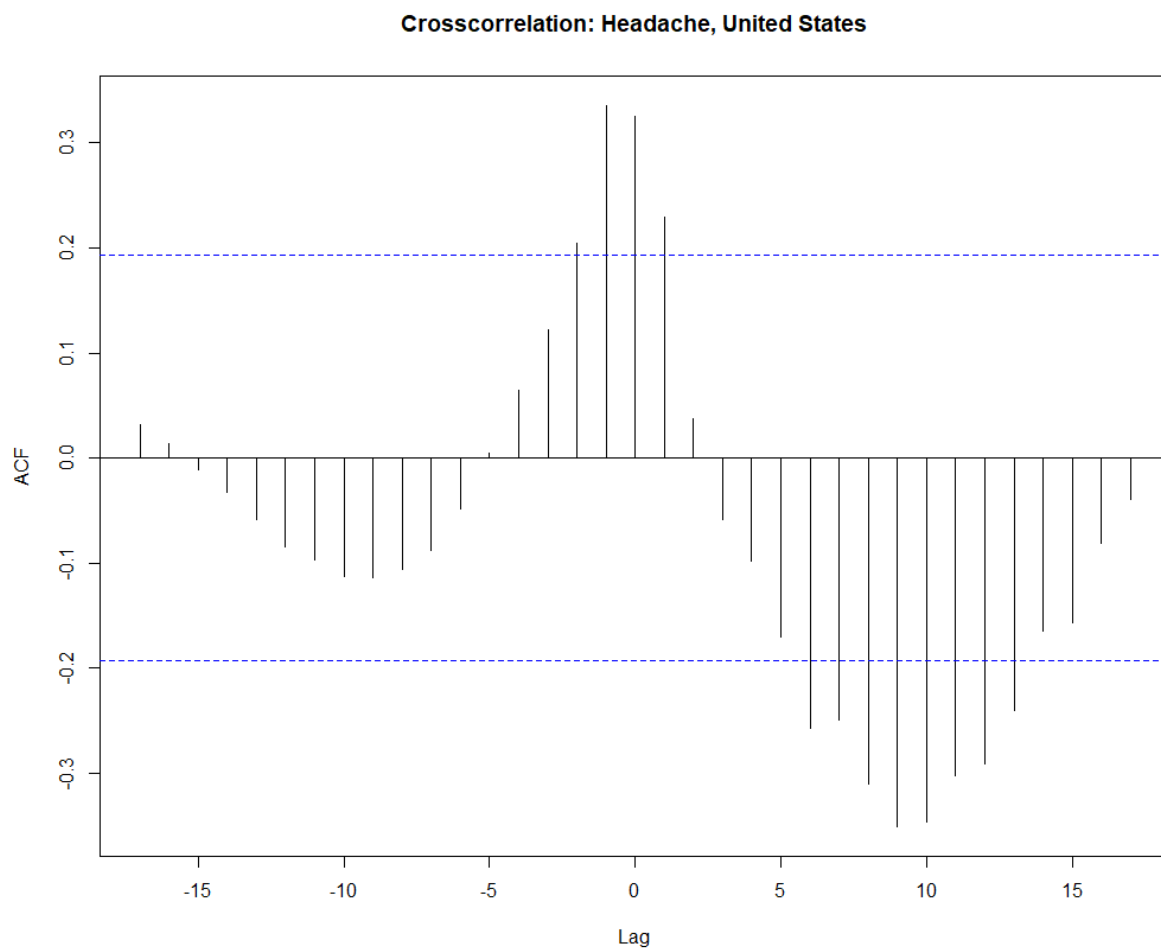


Рисунок 48 График кросскорреляции запросов по теме Головная боль и заболеваемостью в США

Анализ по стране

Лаг в США почти нулевой, но при этом наблюдается схожая с Францией картина первой волны: взрывной рост популярности запросов, несоразмерный заболеваемости во время первой волны.

Благодаря почти нулевому лагу таблице с посчитанной корреляцией можно верить, не проверяя максимума корреляции на графике кросскорреляции.

Индия

Таблица 5 Корреляция заболеваемости и популярности запросов в Индии

WORDKEY	PEARSON	SPEARMAN
COVID	0.088923	0.207482
SMELL	0.868399	0.770105
THROAT	0.380654	0.172382
FEWER	0.858437	0.575025
DYSPNEA	0.634152	0.311559
COUGH	0.694892	0.143767
RHEUM	0.712237	0.392717
HEADACHE	0.775421	0.526135

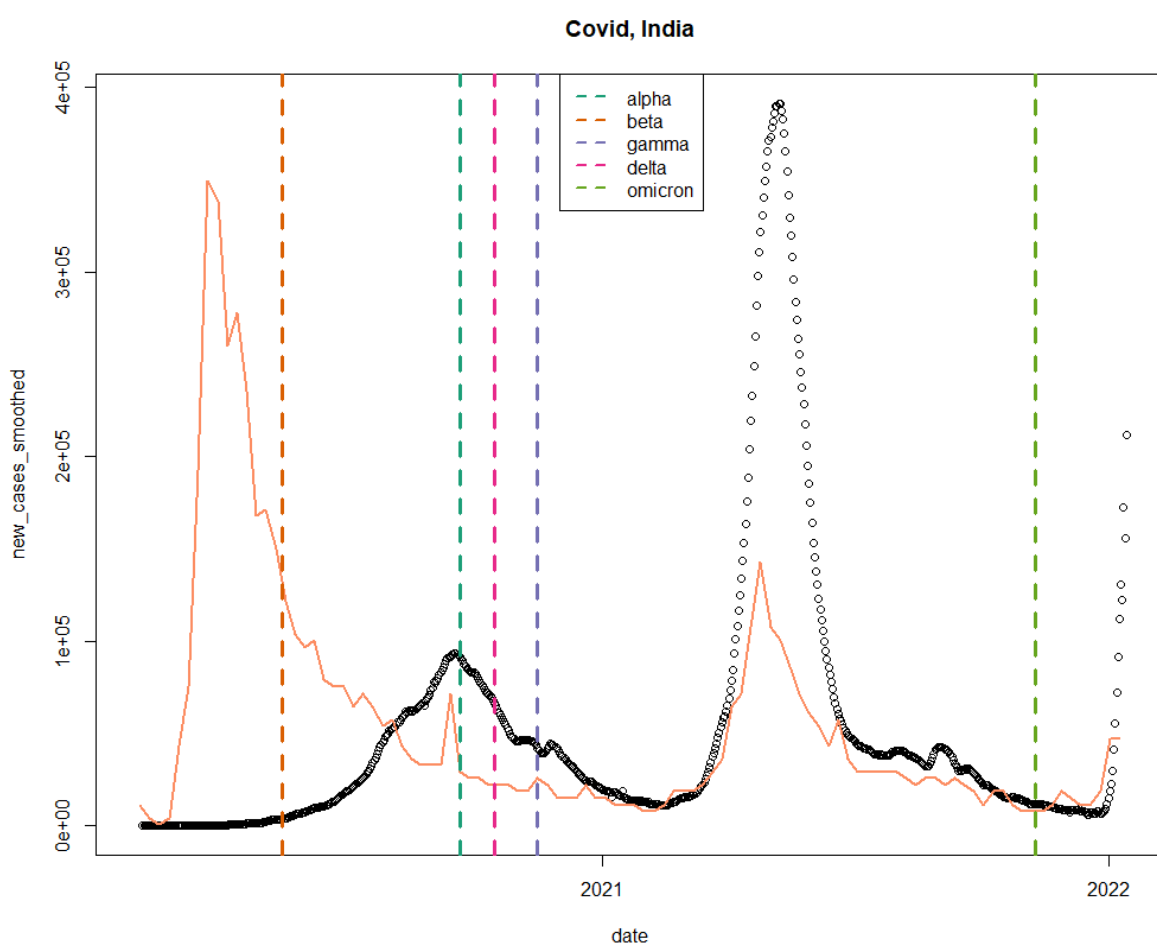


Рисунок 49 Графики популярности запросов по теме COVID-19 и подтвержденных случаев заболевания в Индии

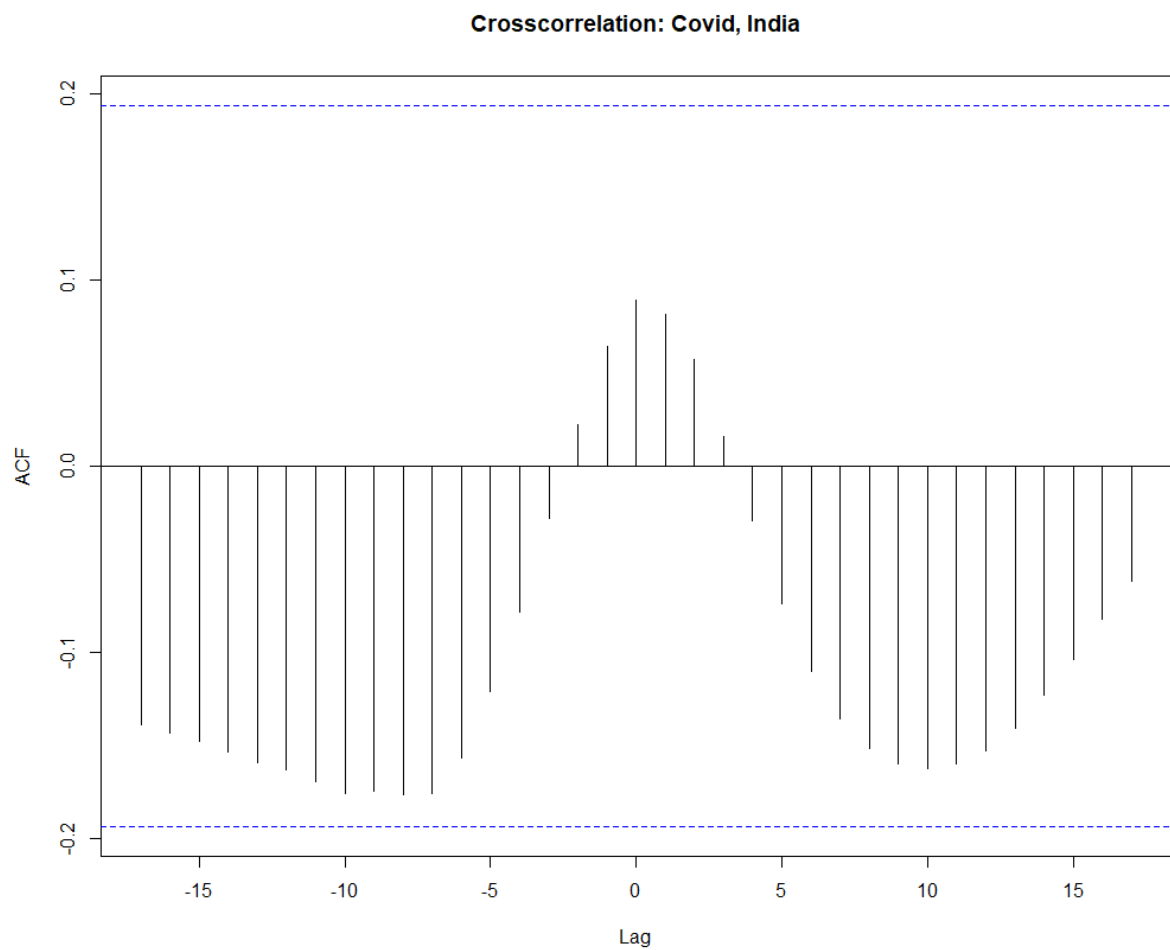


Рисунок 50 График кросскорреляции запросов по теме COVID-19 и заболеваемостью в Индии

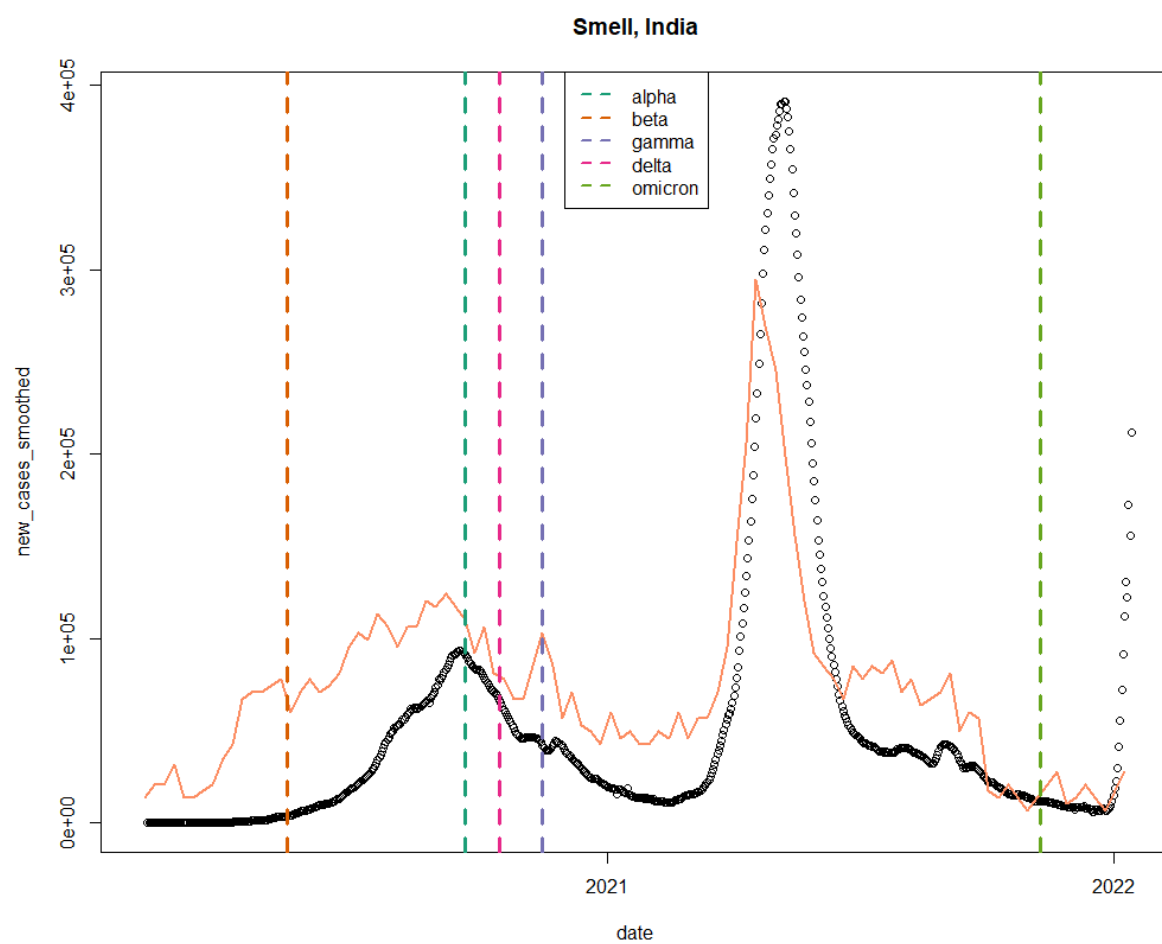


Рисунок 51 Графики популярности запросов по теме Обоняние и подтвержденных случаев заболевания в Индии

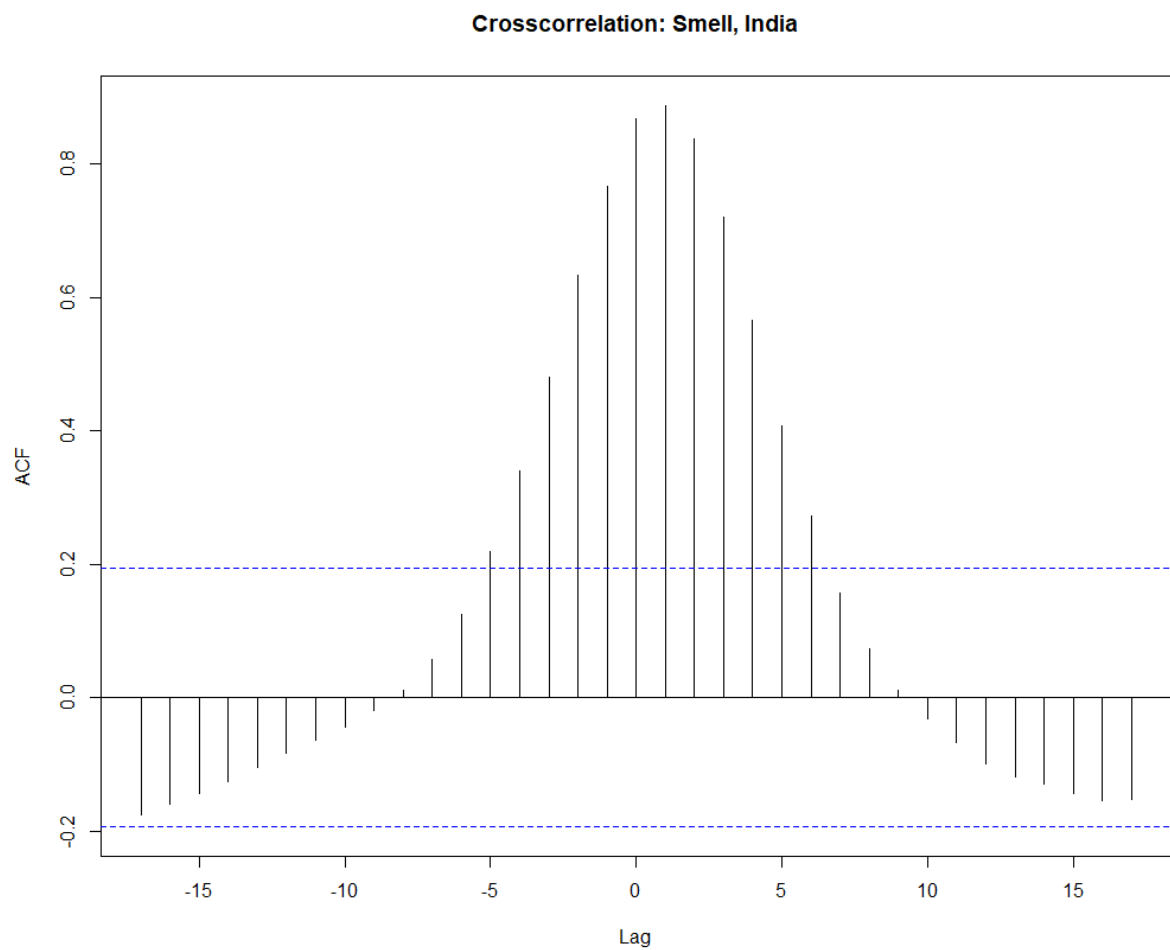


Рисунок 52 График кросскорреляции запросов по теме Обоняние и заболеваемостью в Индии

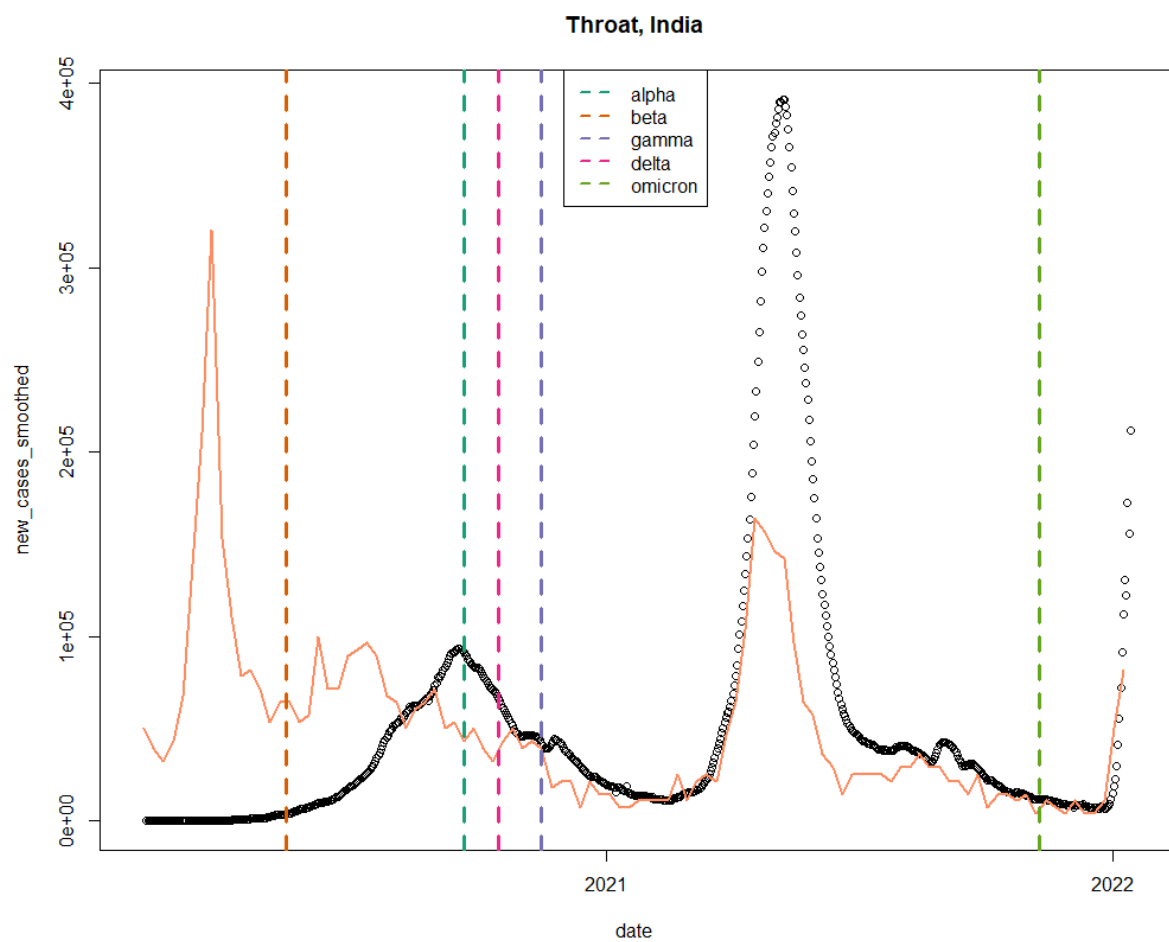


Рисунок 53 Графики популярности запросов по теме Больное горло и подтвержденных случаев заболевания в Индии

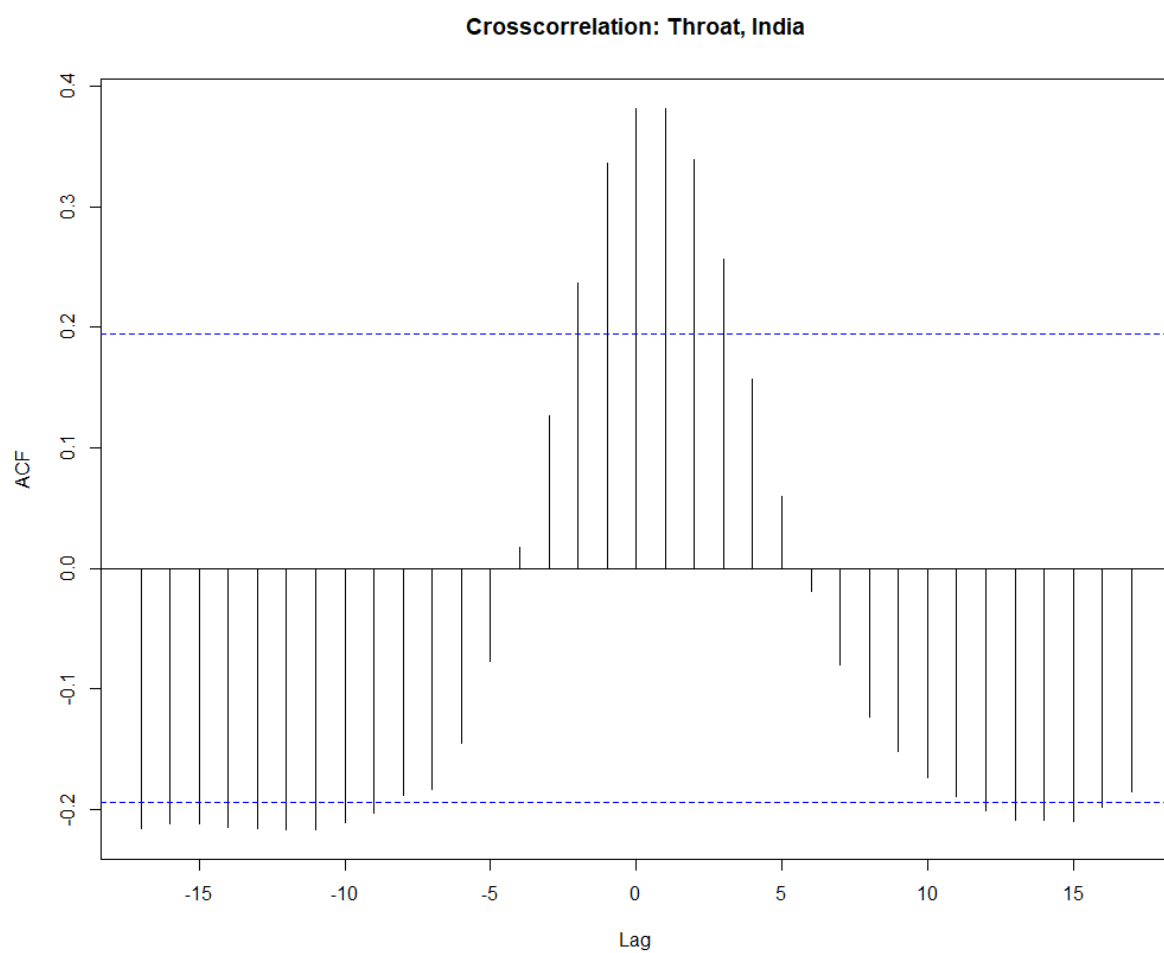


Рисунок 54 График кросскорреляции запросов по теме Боль в горле и заболеваемостью в Индии

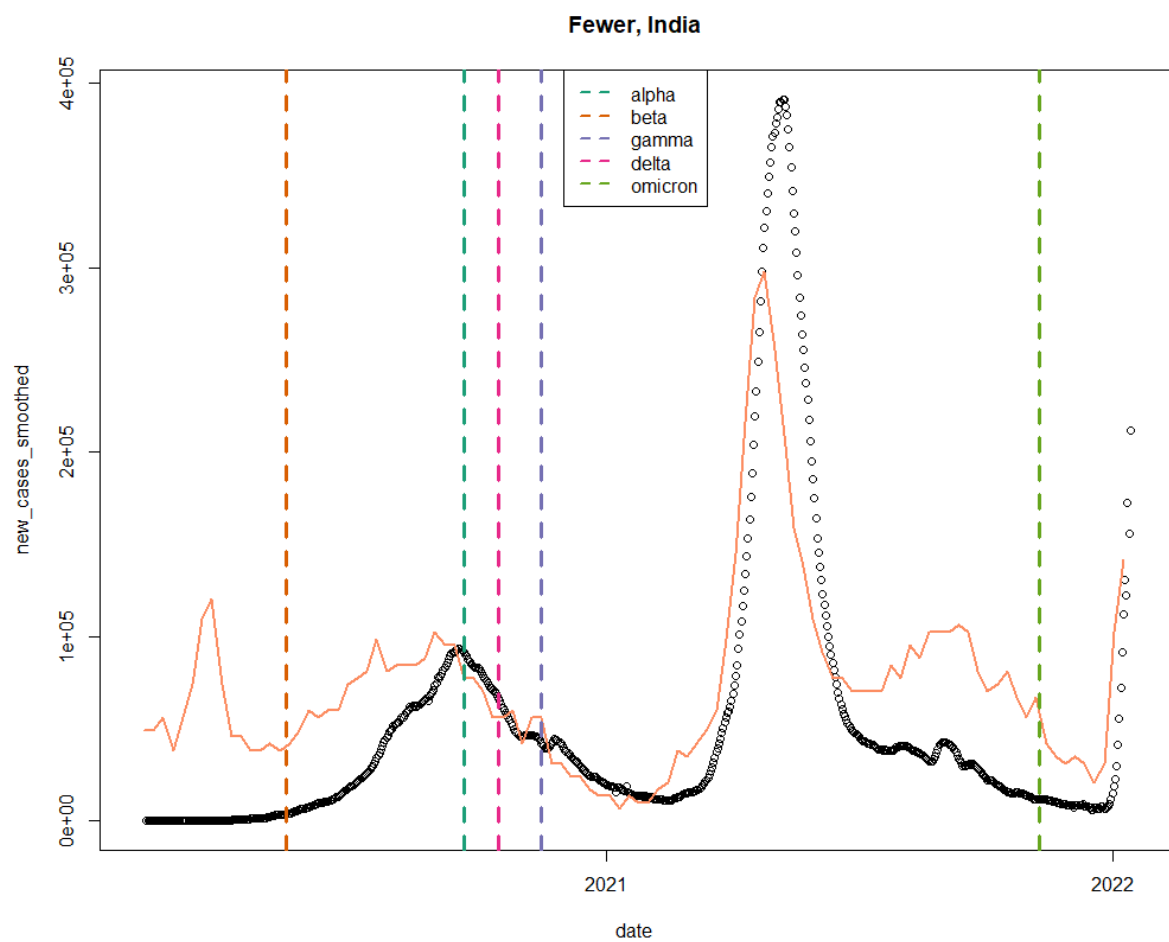


Рисунок 55 Графики популярности запросов по теме Лихорадка и подтвержденных случаев заболевания в Индии

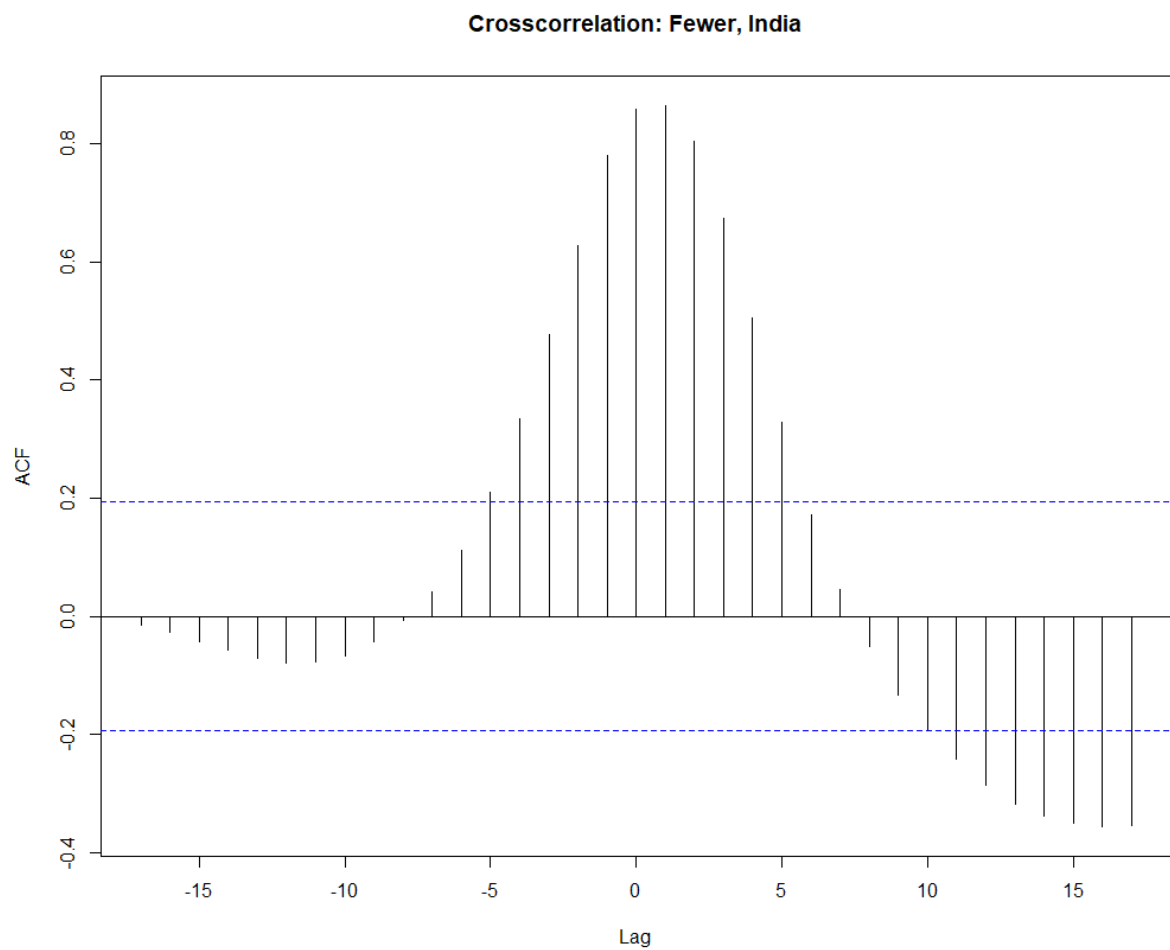


Рисунок 56 График кросскорреляции запросов по теме Лихорадка и заболеваемостью в Индии

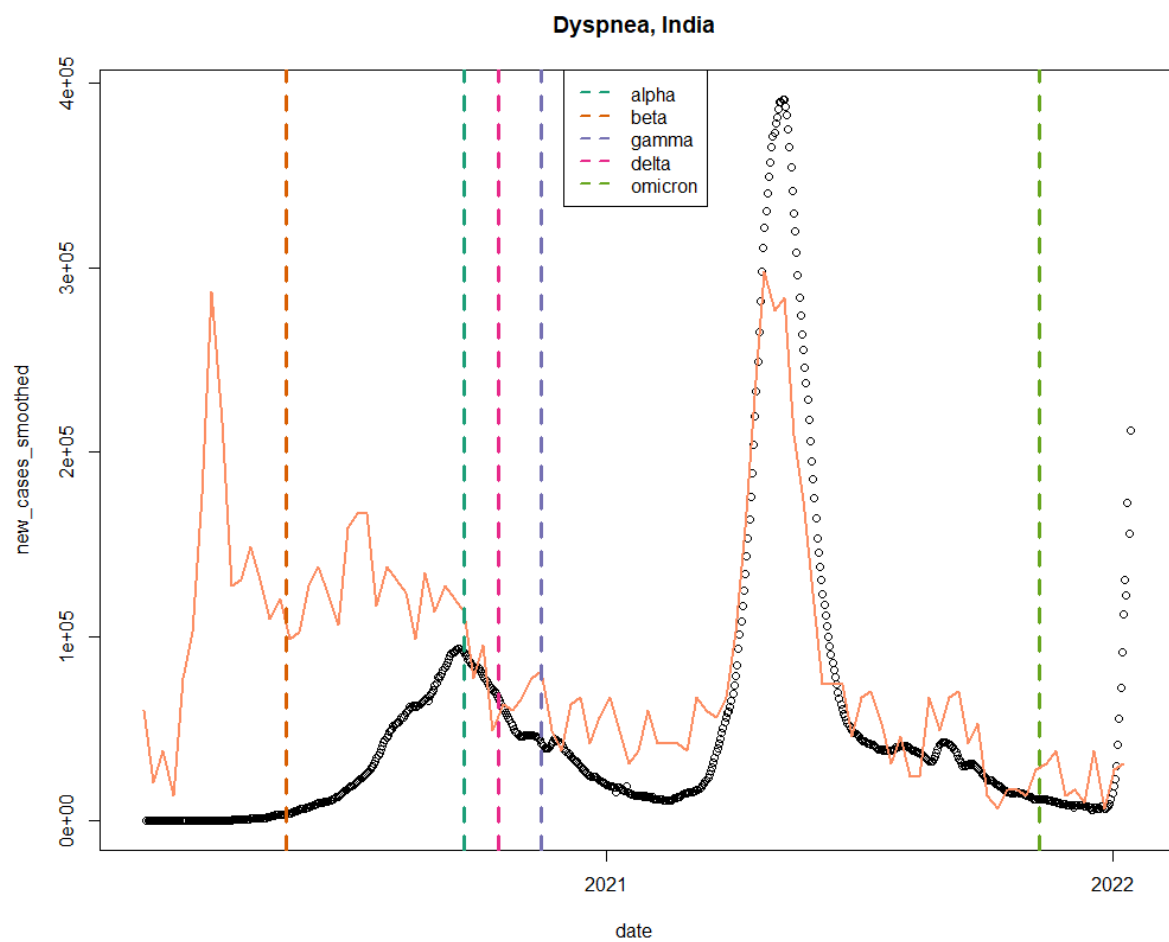


Рисунок 57 Графики популярности запросов по теме Одышка и подтвержденных случаев заболевания в Индии

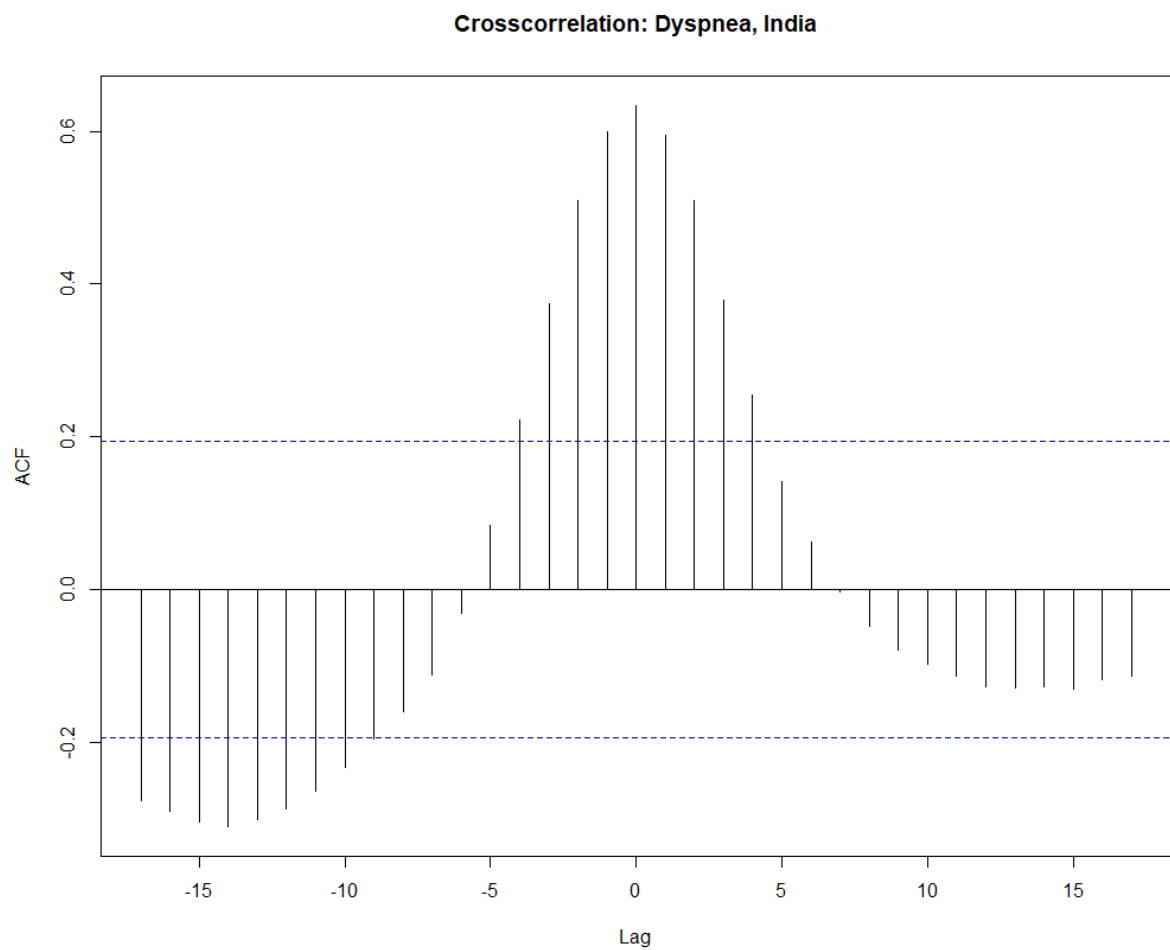


Рисунок 58 График кросскорреляции запросов по теме Одышка и заболеваемостью в Индии

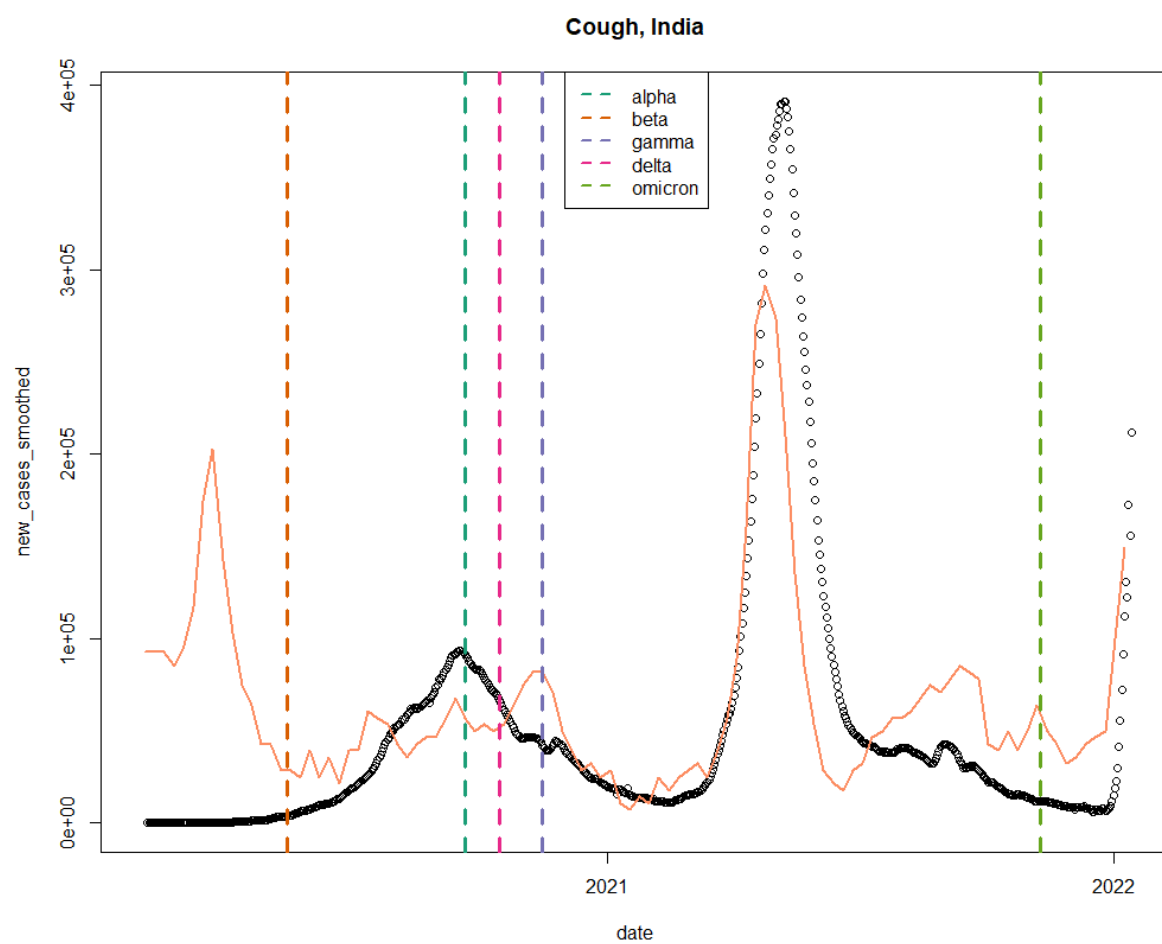


Рисунок 59 Графики популярности запросов по теме Кашель и подтвержденных случаев заболевания в Индии

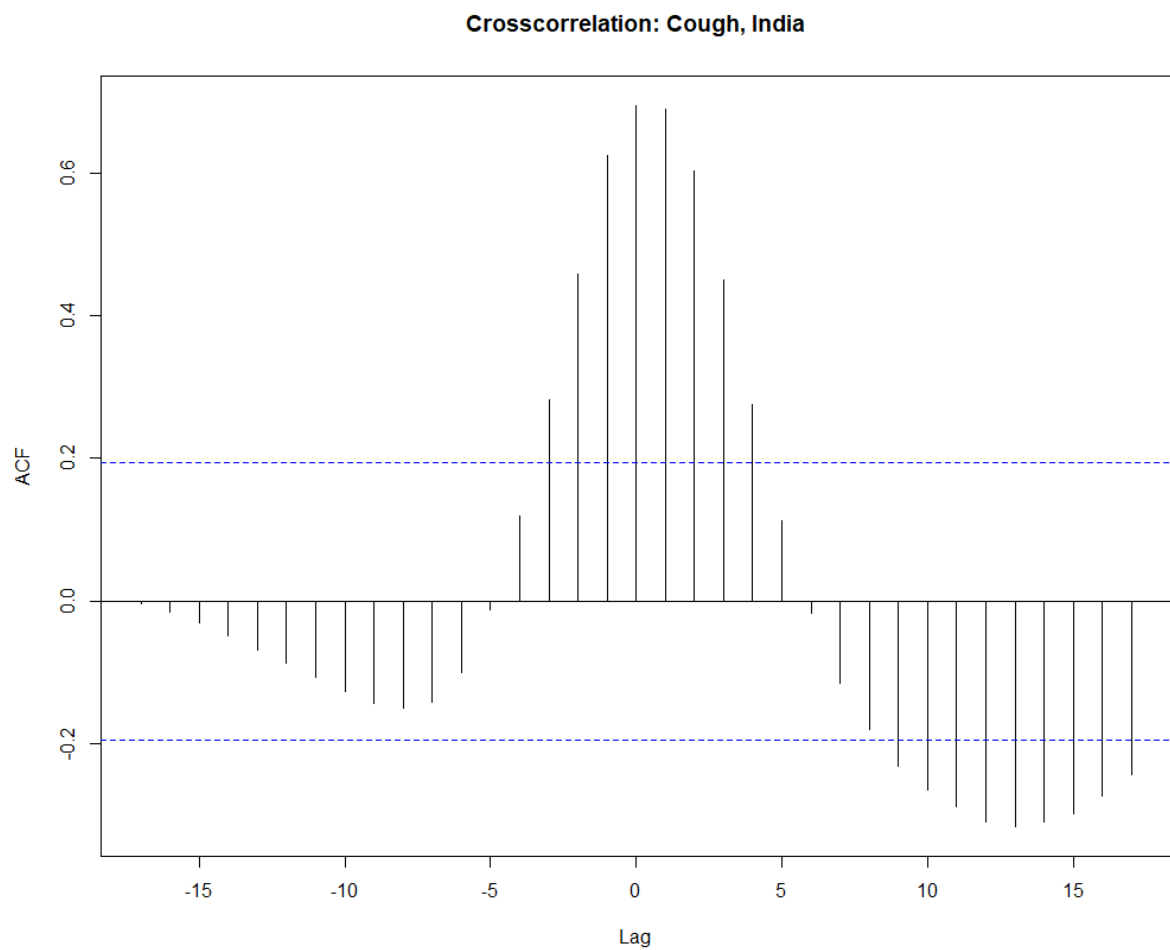


Рисунок 60 График кросскорреляции запросов по теме Кашель и заболеваемостью в Индии

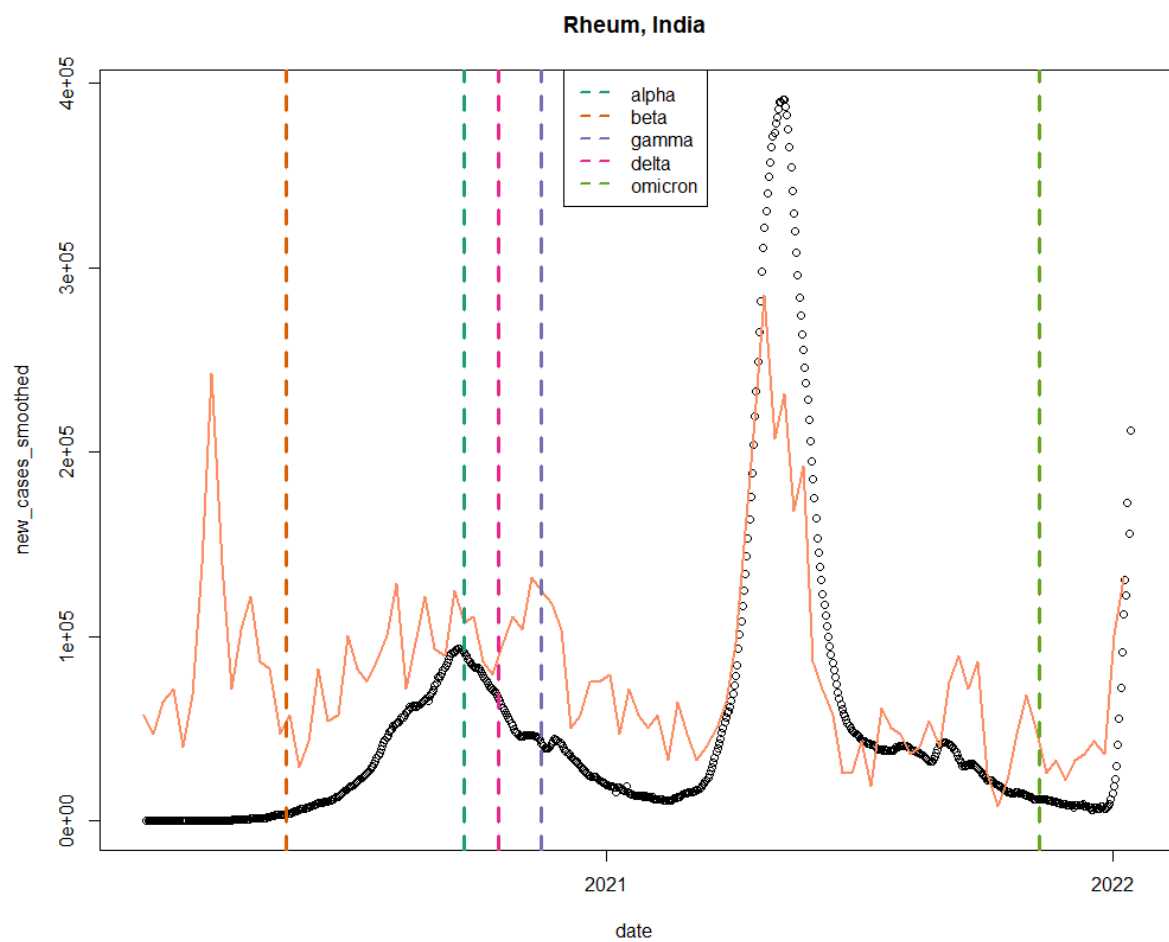


Рисунок 61 Графики популярности запросов по теме Насморк и подтвержденных случаев заболевания в Индии

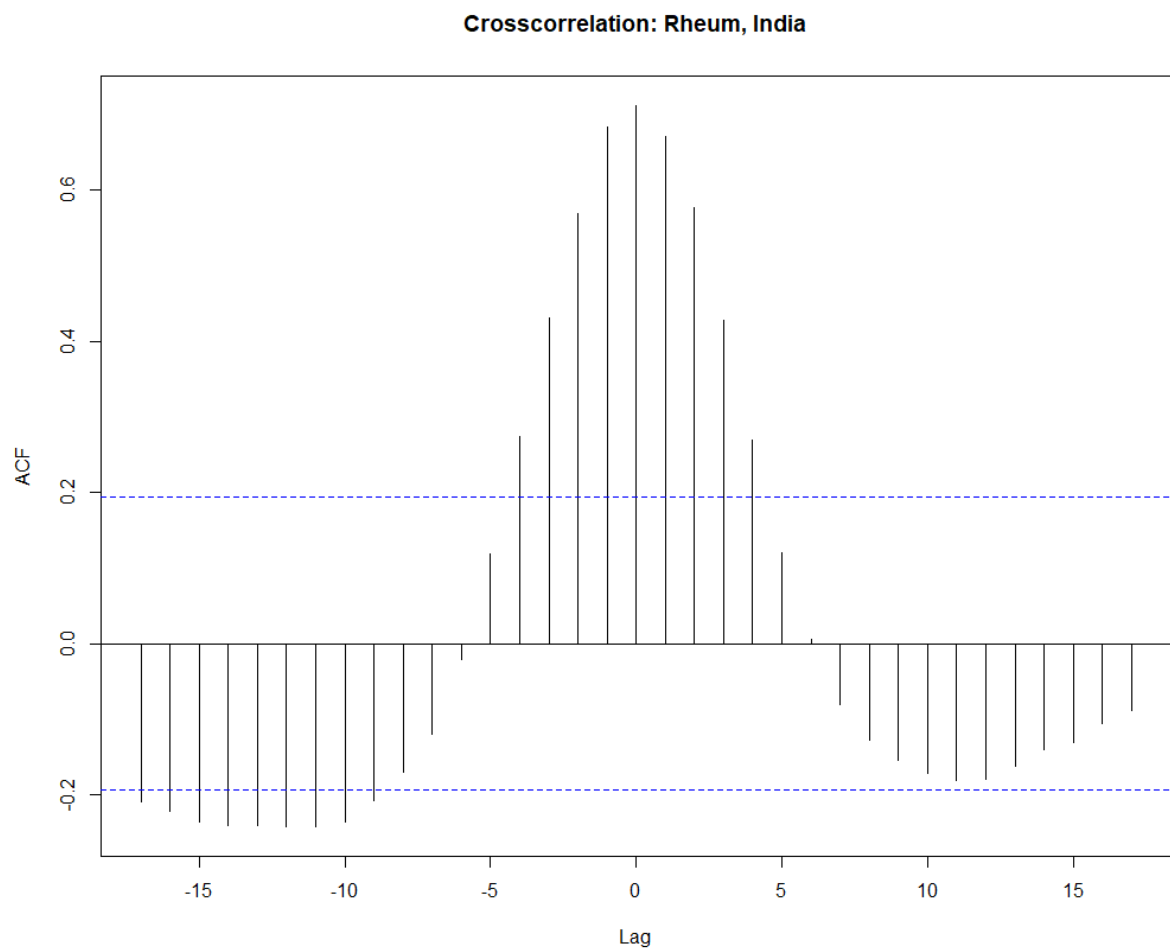


Рисунок 62 График кросскорреляции запросов по теме Насморк и заболеваемостью в Индии

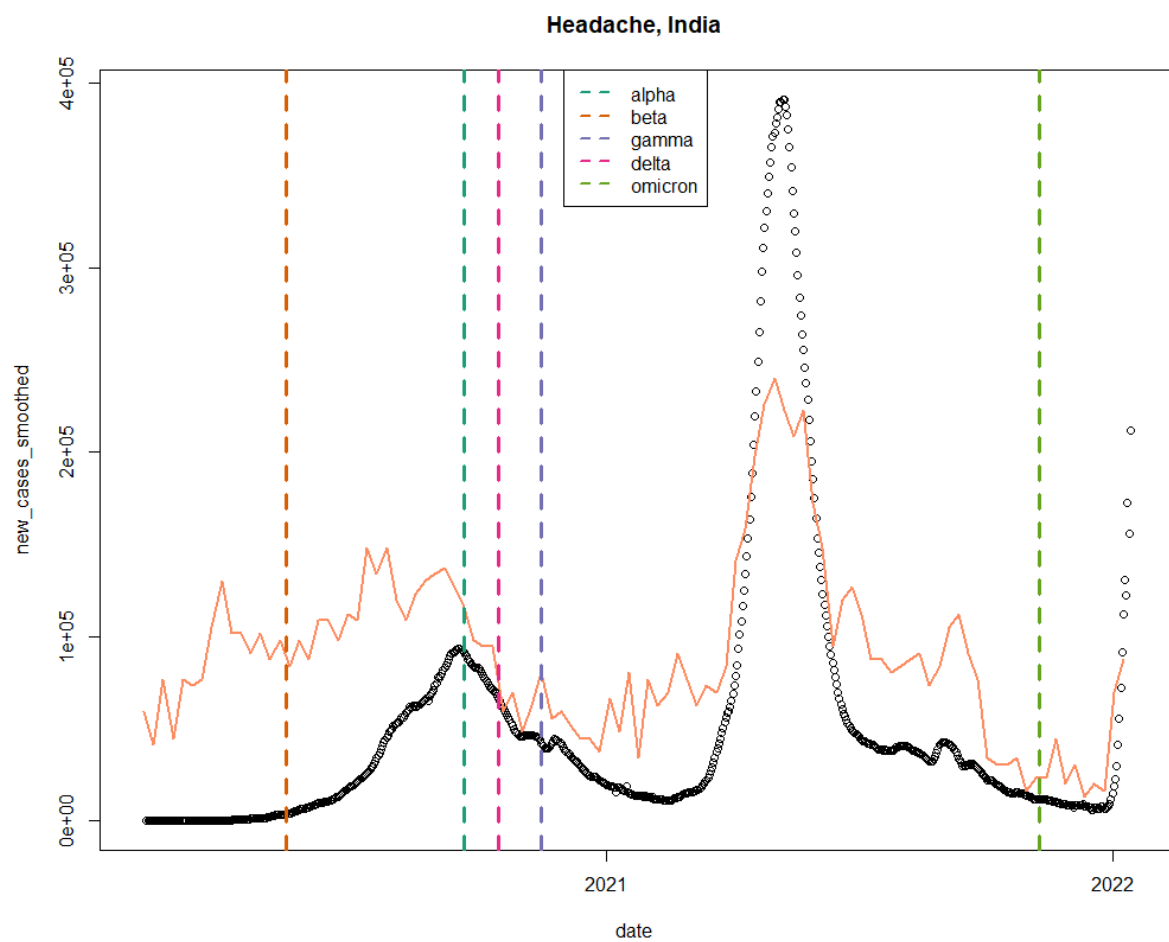


Рисунок 63 Графики популярности запросов по теме Головная боль и подтвержденных случаев заболевания в Индии

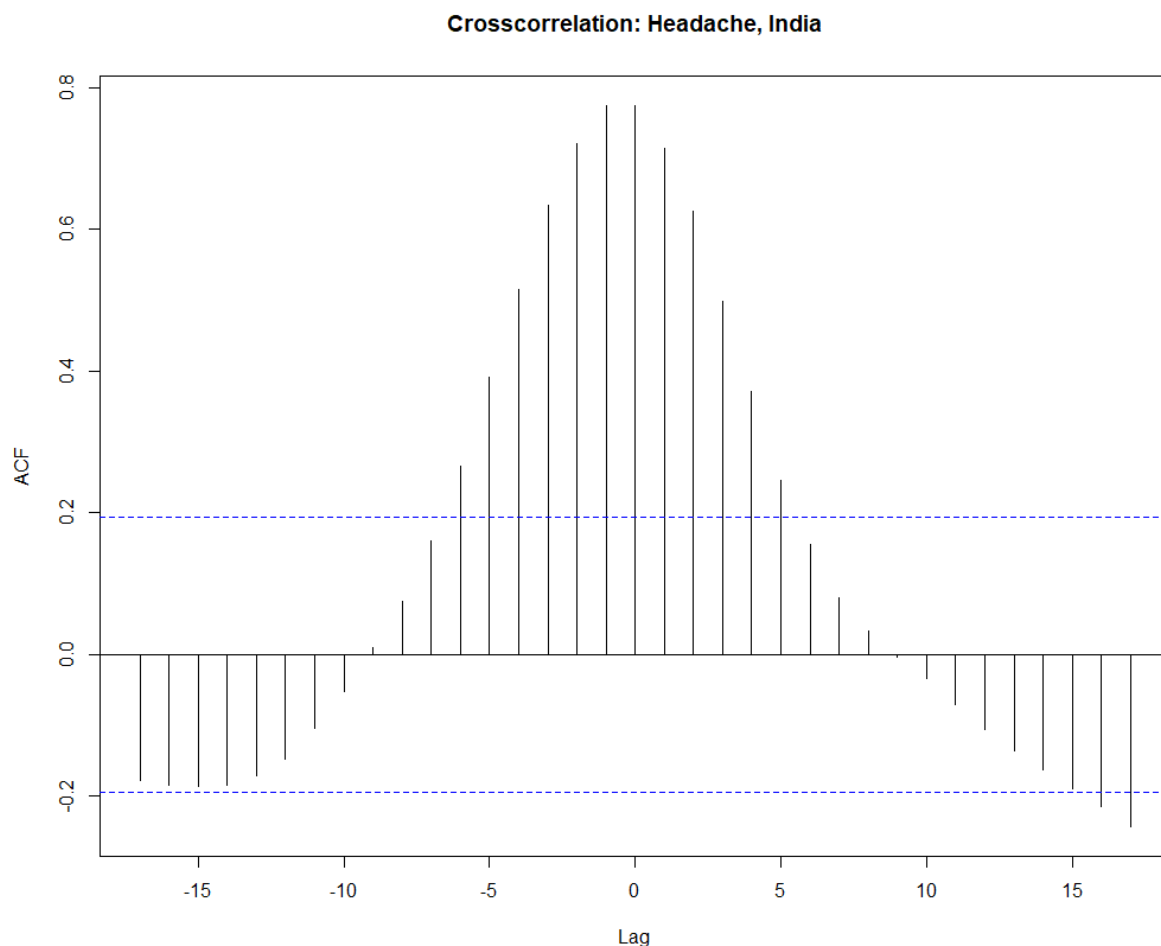


Рисунок 64 График кросскорреляции запросов по теме Головная боль и заболеваемостью в Индии

Анализ по стране

Индия – самая показательная страна для исследования. Корреляция почти по всем запросам > 0.6 (кроме темы «Боль в горле»). Это удивительные значения корреляции.

По некоторым графикам можно сказать, что в стране была совсем не учтена первая волна заболеваемости. Что интересно, в эту волну рост запросов по теме «Обоняние» не сильно выражен. Это может говорить о том, что у первоначальной вариации вируса, который попал в страну потеря обоняния не была таким частым симптомом, как у дельты, которая как раз впервые была зарегистрирована на территории Индии.

Великобритания

Таблица 6 Корреляция заболеваемости и популярности запросов в Великобритании

WORDKEY	PEARSON	SPEARMAN
COVID	-0.10118	-0.04721
SMELL	0.021007	0.299734
THROAT	0.029718	0.306191
FEWER	0.024518	0.398593
DYSPNEA	-0.15475	0.016459
COUGH	0.18801	0.478509
RHEUM	-0.07954	0.228139
HEADACHE	0.002881	0.102502

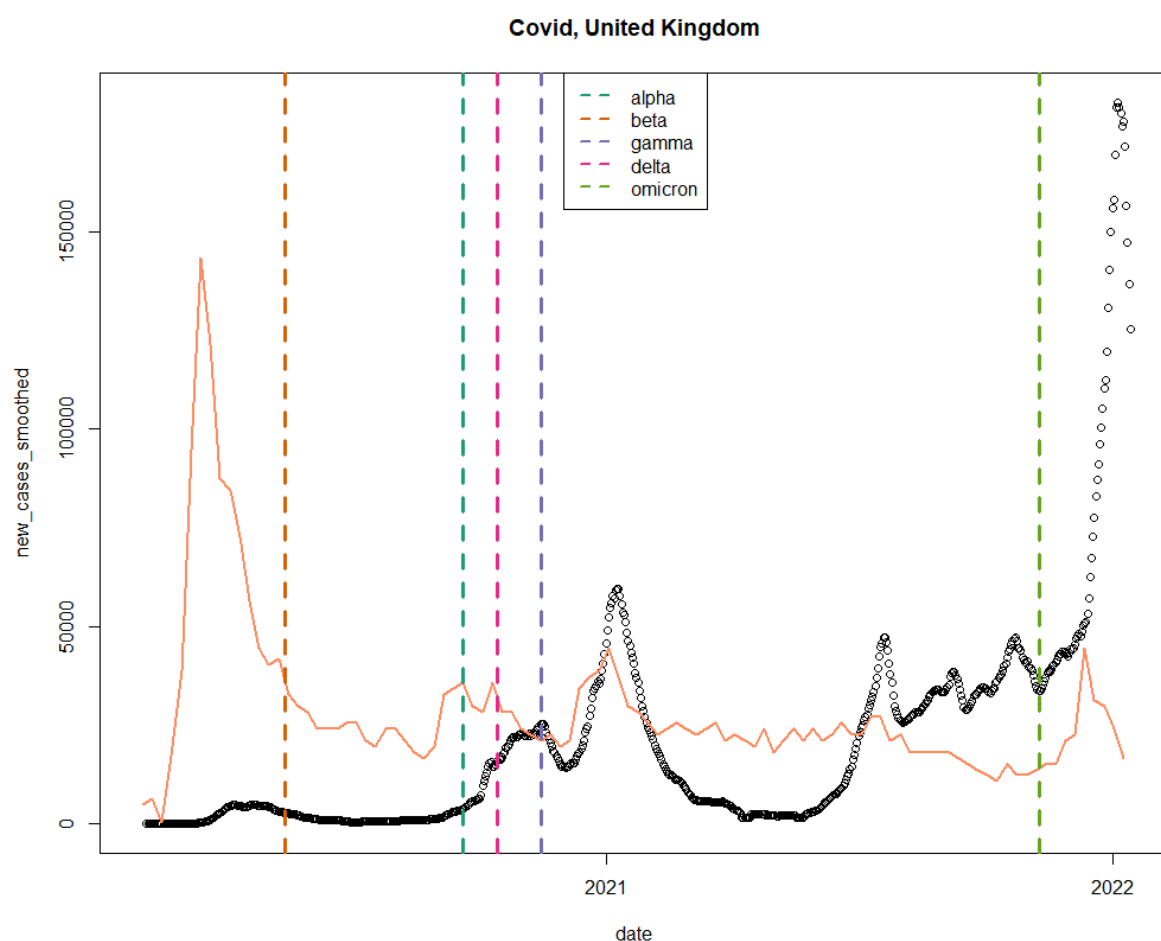


Рисунок 65 Графики популярности запросов по теме COVID-19 и подтвержденных случаев заболевания в Великобритании

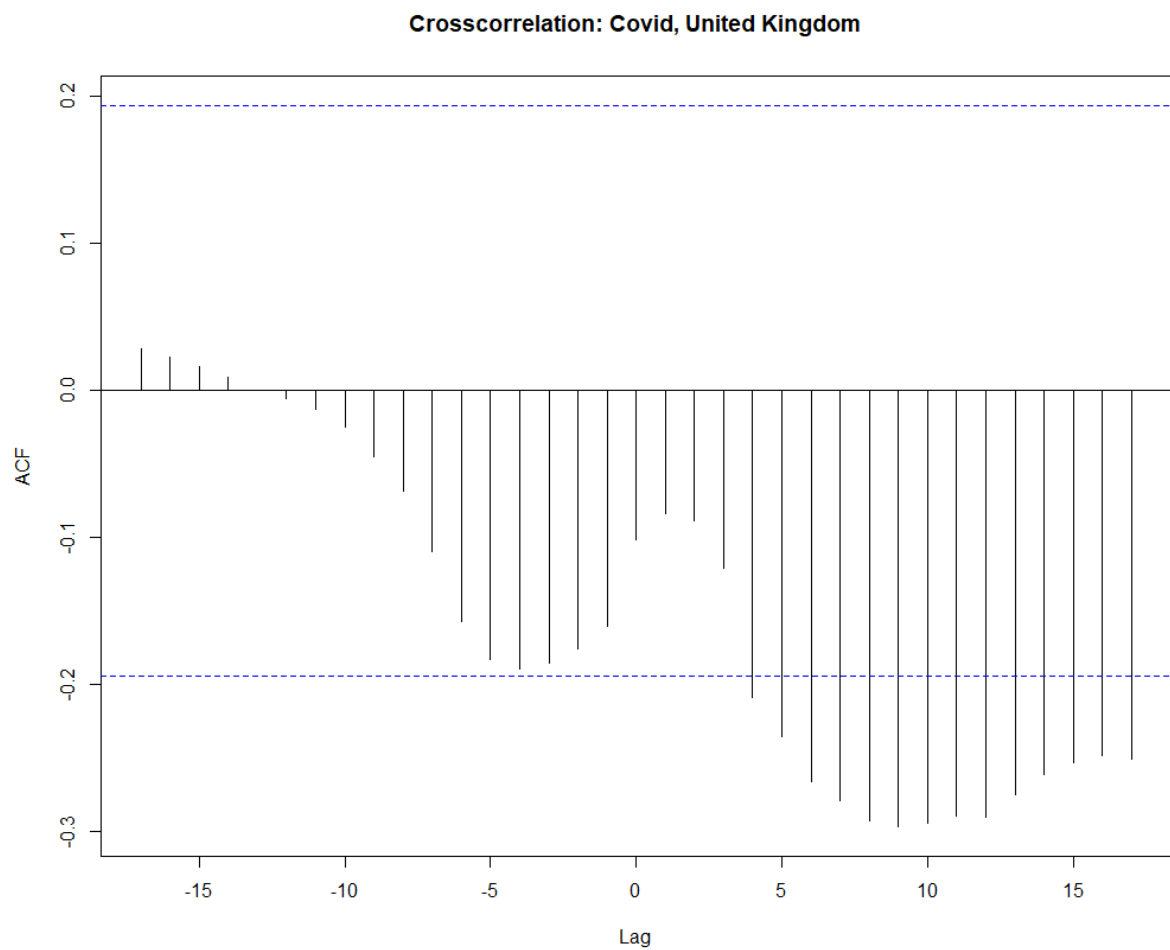


Рисунок 66 График кросскорреляции запросов по теме COVID-19 и заболеваемостью в Великобритании

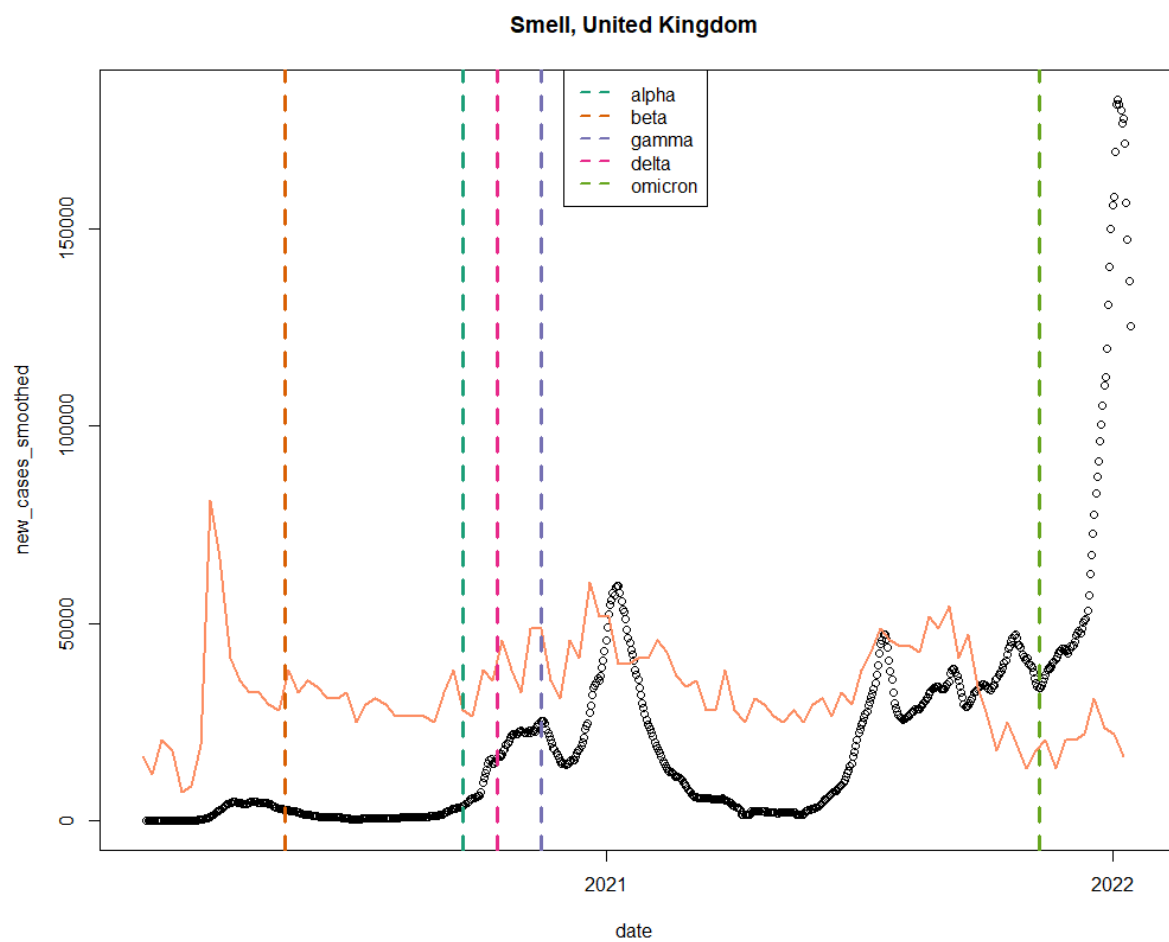


Рисунок 67 Графики популярности запросов по теме Обоняние и подтвержденных случаев заболевания в Великобритании

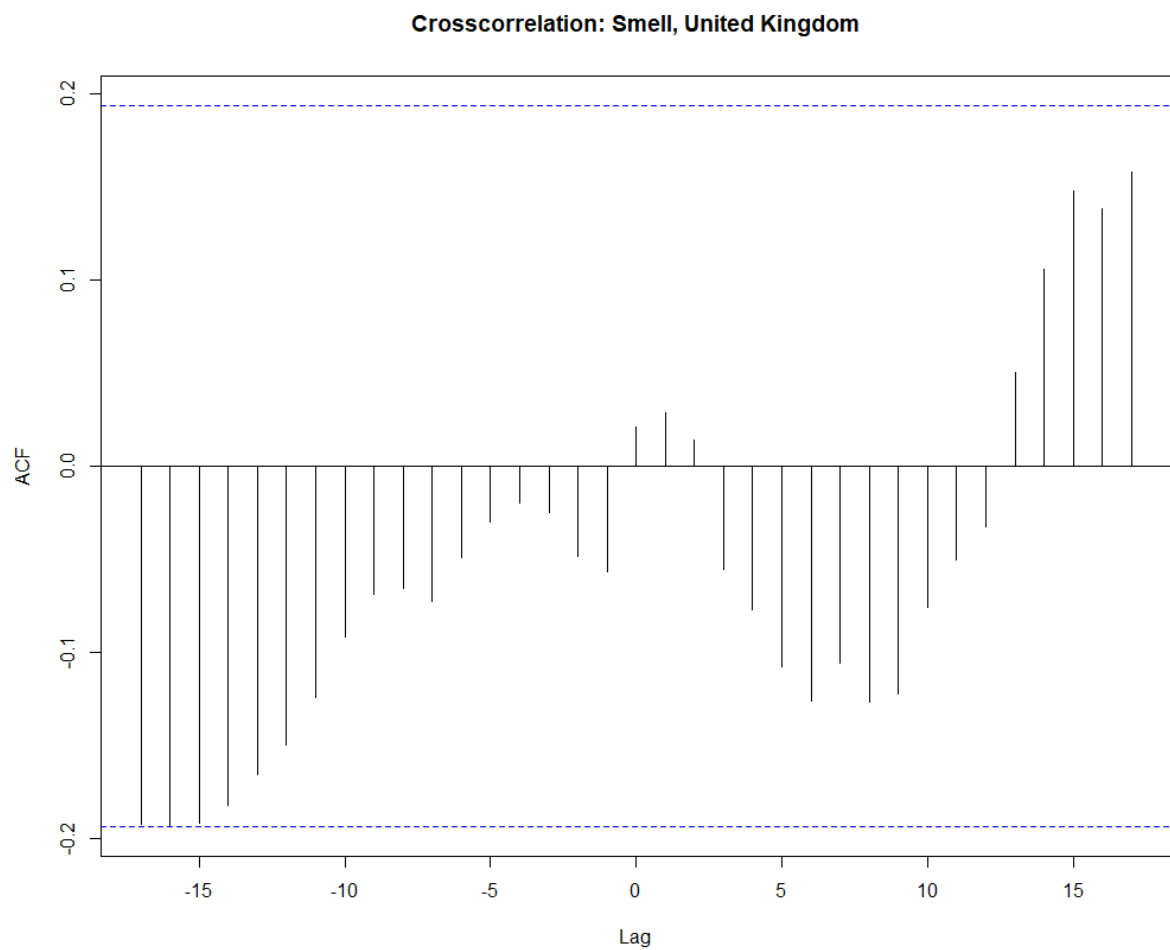


Рисунок 68 График кросскорреляции запросов по теме Обоняние и заболеваемостью в Великобритании

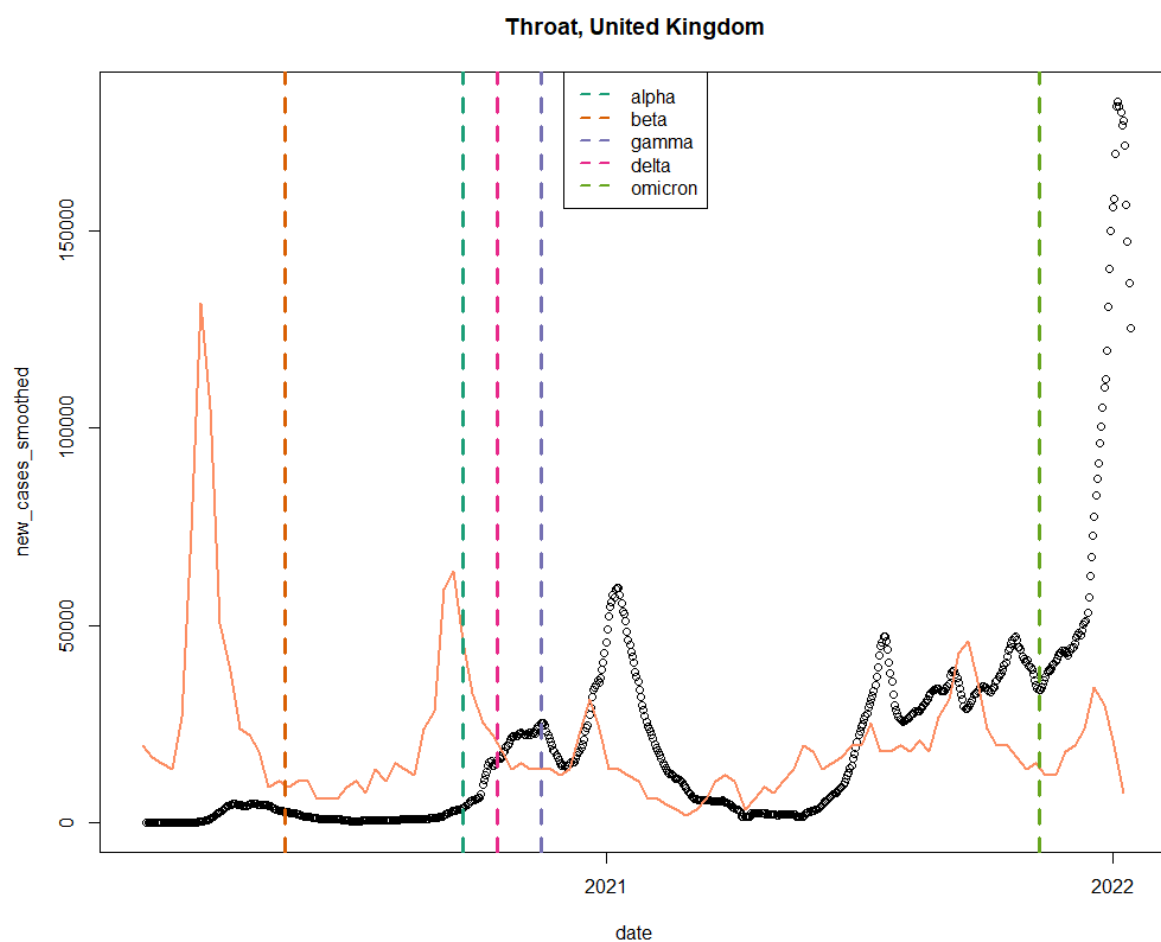


Рисунок 69 Графики популярности запросов по теме Больное горло и подтвержденных случаев заболевания в Великобритании

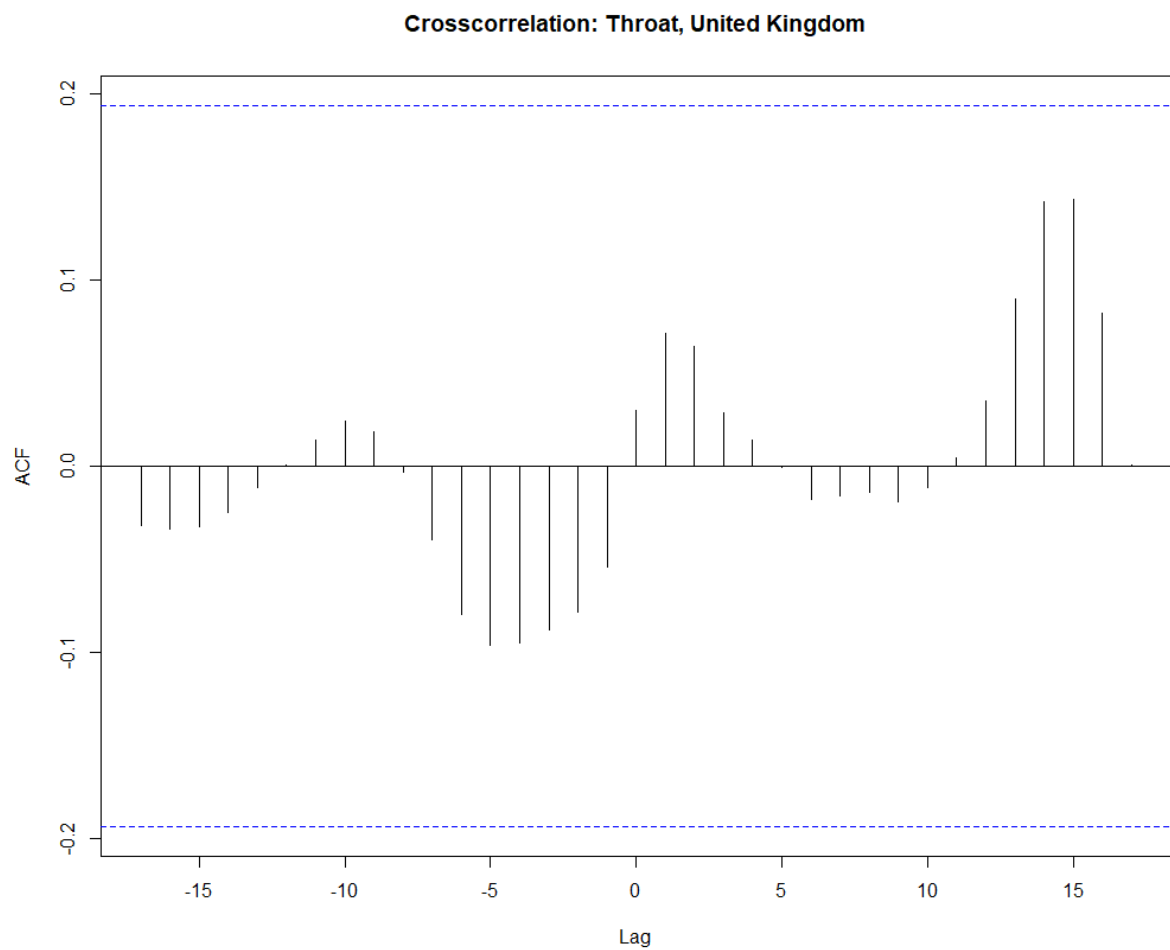


Рисунок 70 График кросскорреляции запросов по теме Больное горло и заболеваемостью в Великобритании

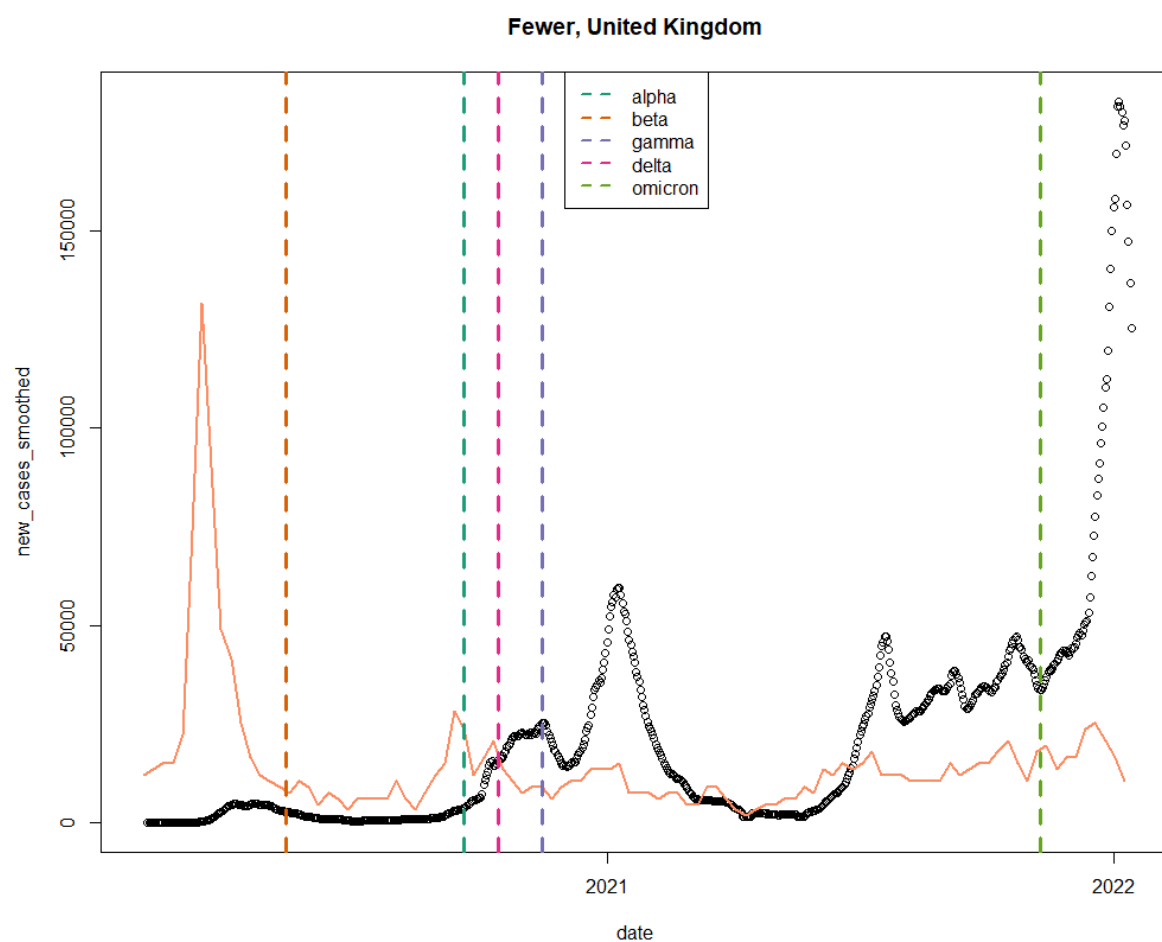


Рисунок 71 Графики популярности запросов по теме Лихорадка и подтвержденных случаев заболевания в Великобритании

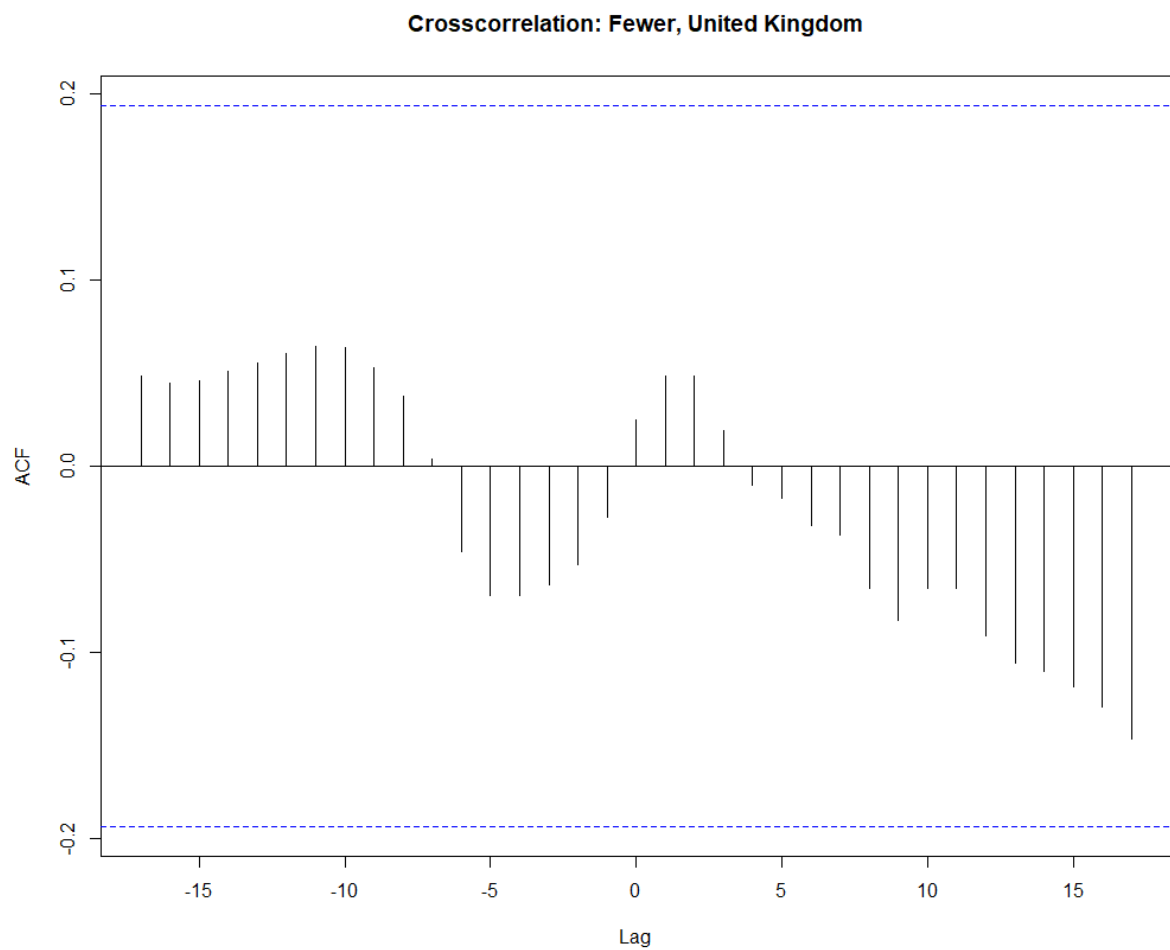


Рисунок 72 График кросскорреляции запросов по теме Лихорадка и заболеваемостью в Великобритании

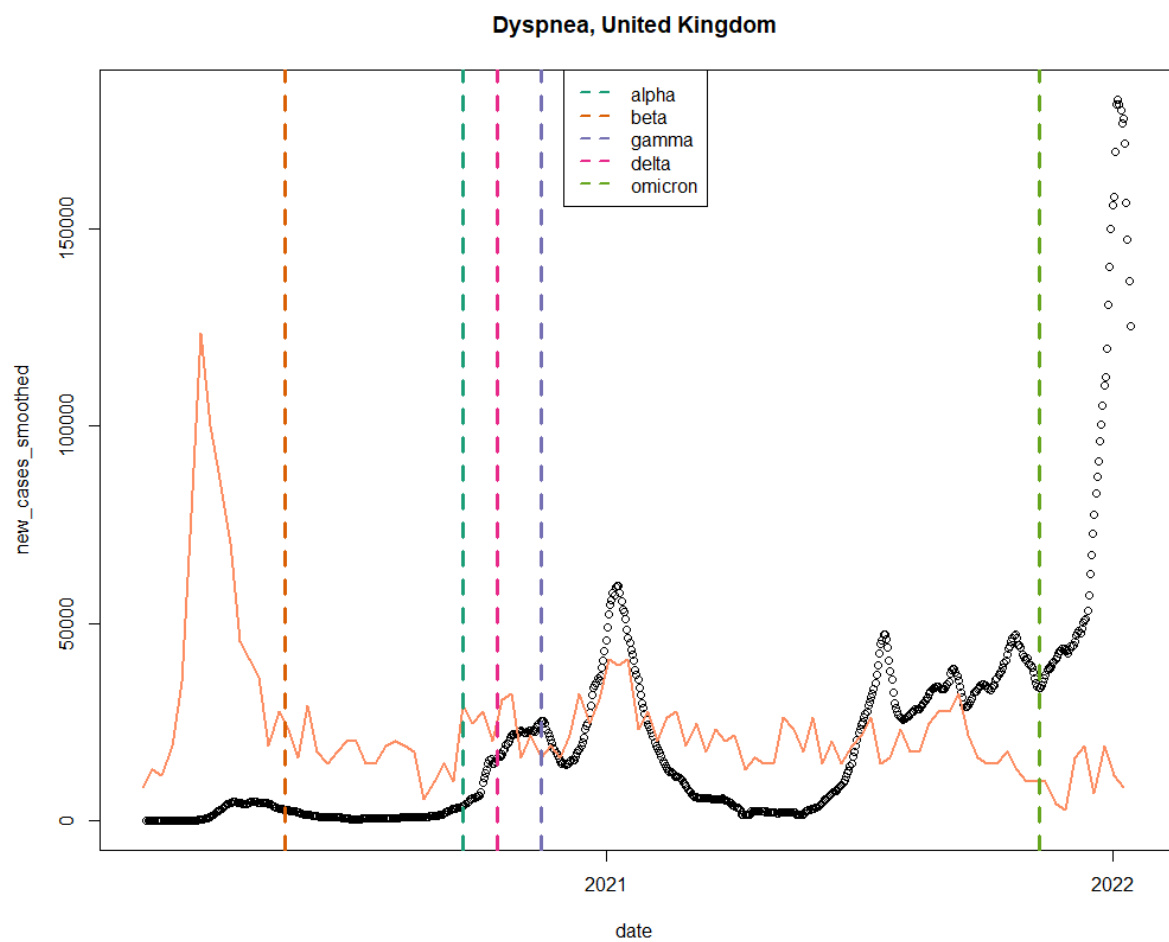


Рисунок 73 Графики популярности запросов по теме Одышка и подтвержденных случаев заболевания в Великобритании

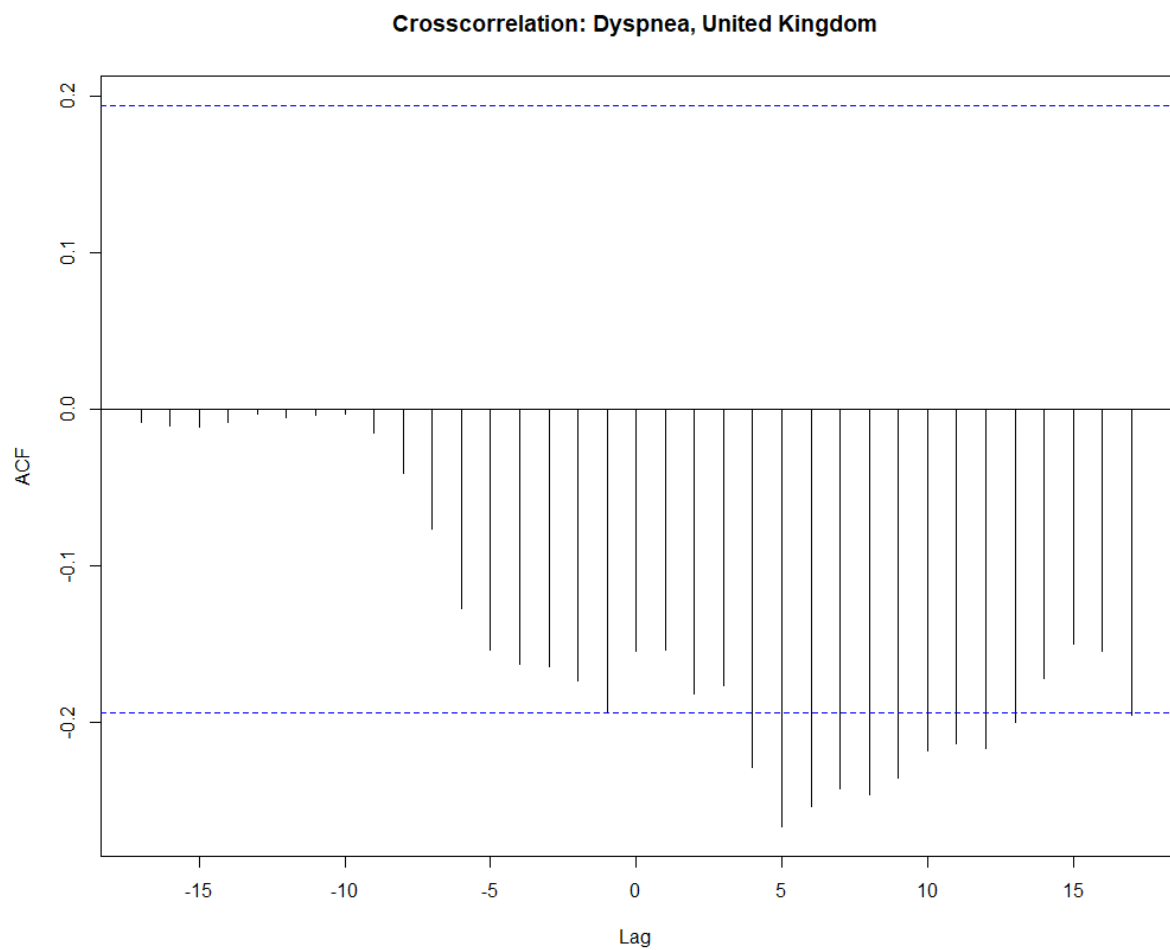


Рисунок 74 График кросскорреляции запросов по теме Одышка и заболеваемостью в Великобритании

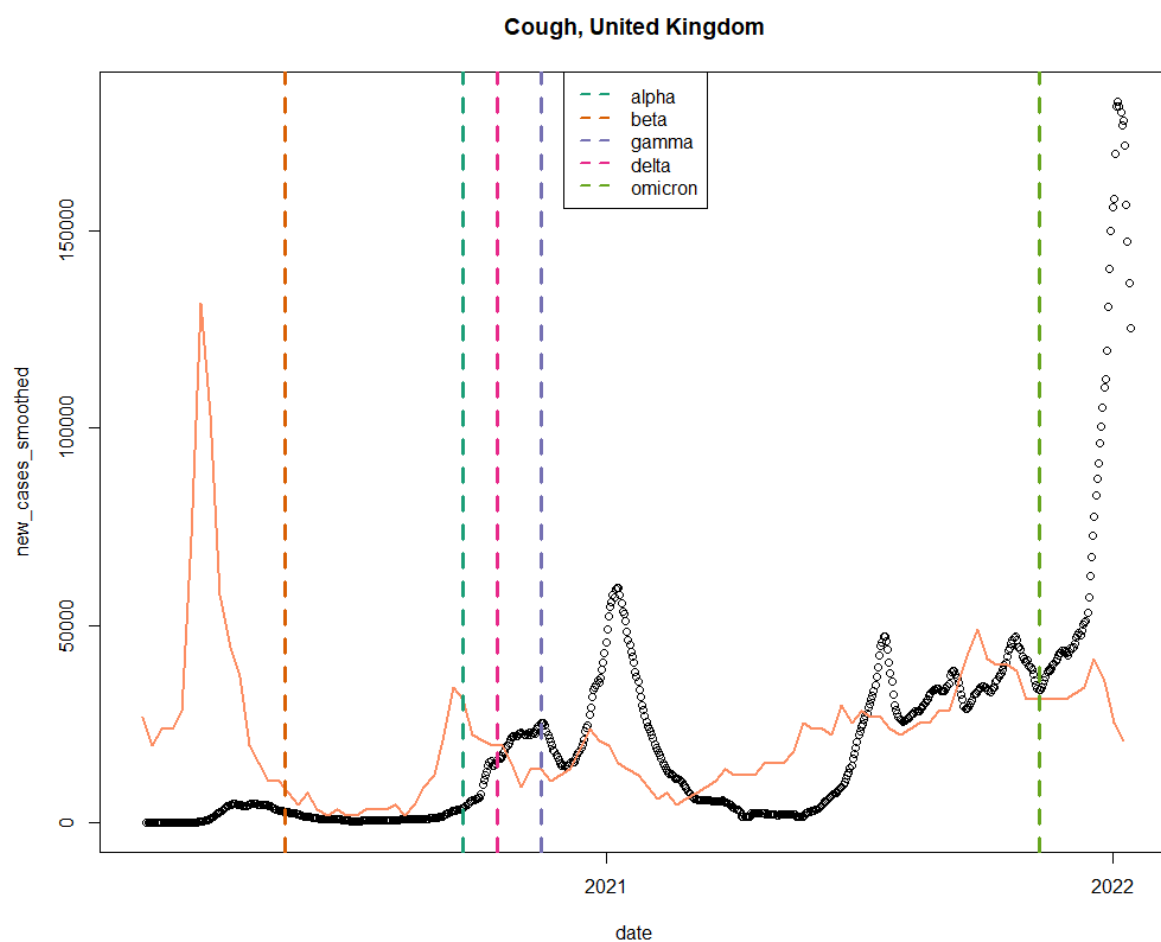


Рисунок 75 Графики популярности запросов по теме Кашель и подтвержденных случаев заболевания в Великобритании

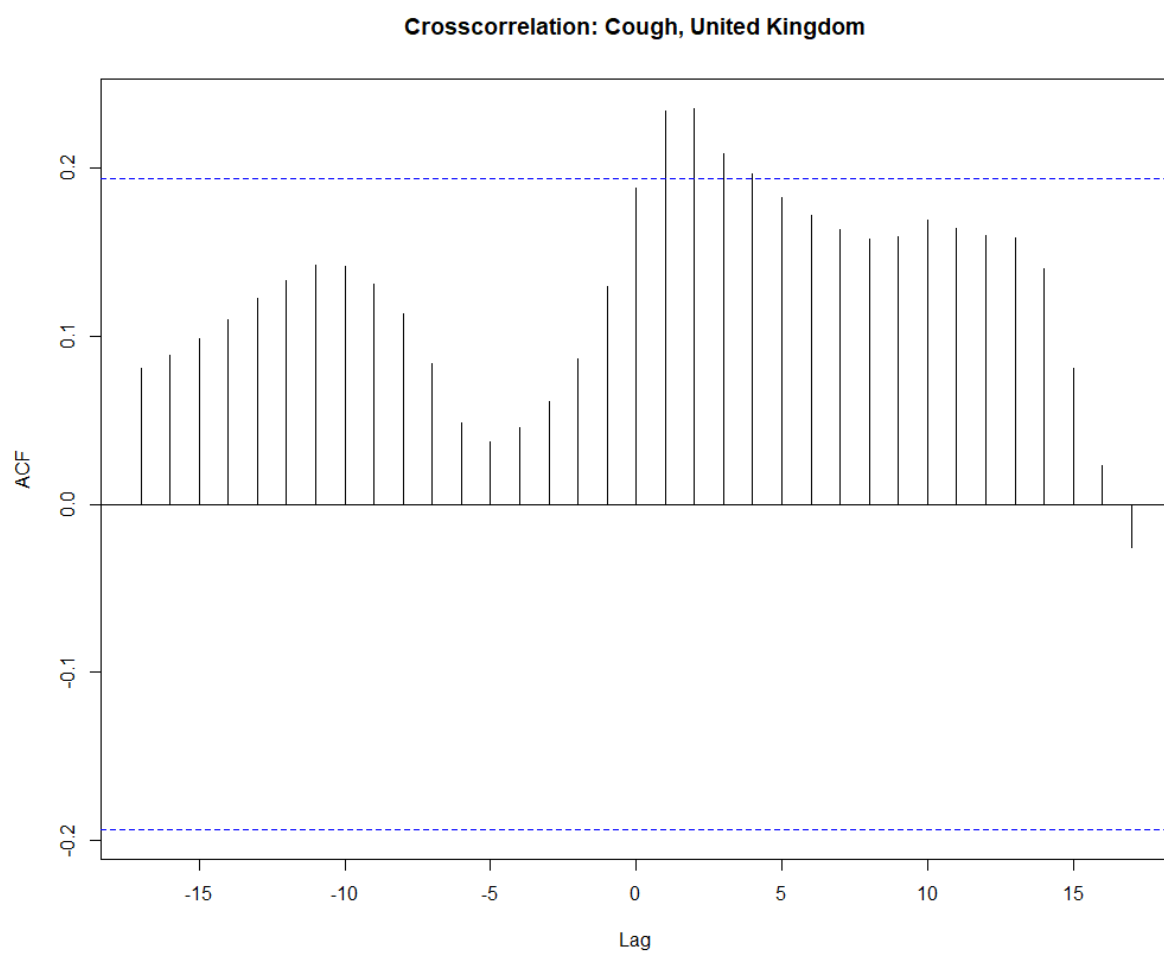


Рисунок 76 График кросскорреляции запросов по теме Кашель и заболеваемостью в Великобритании

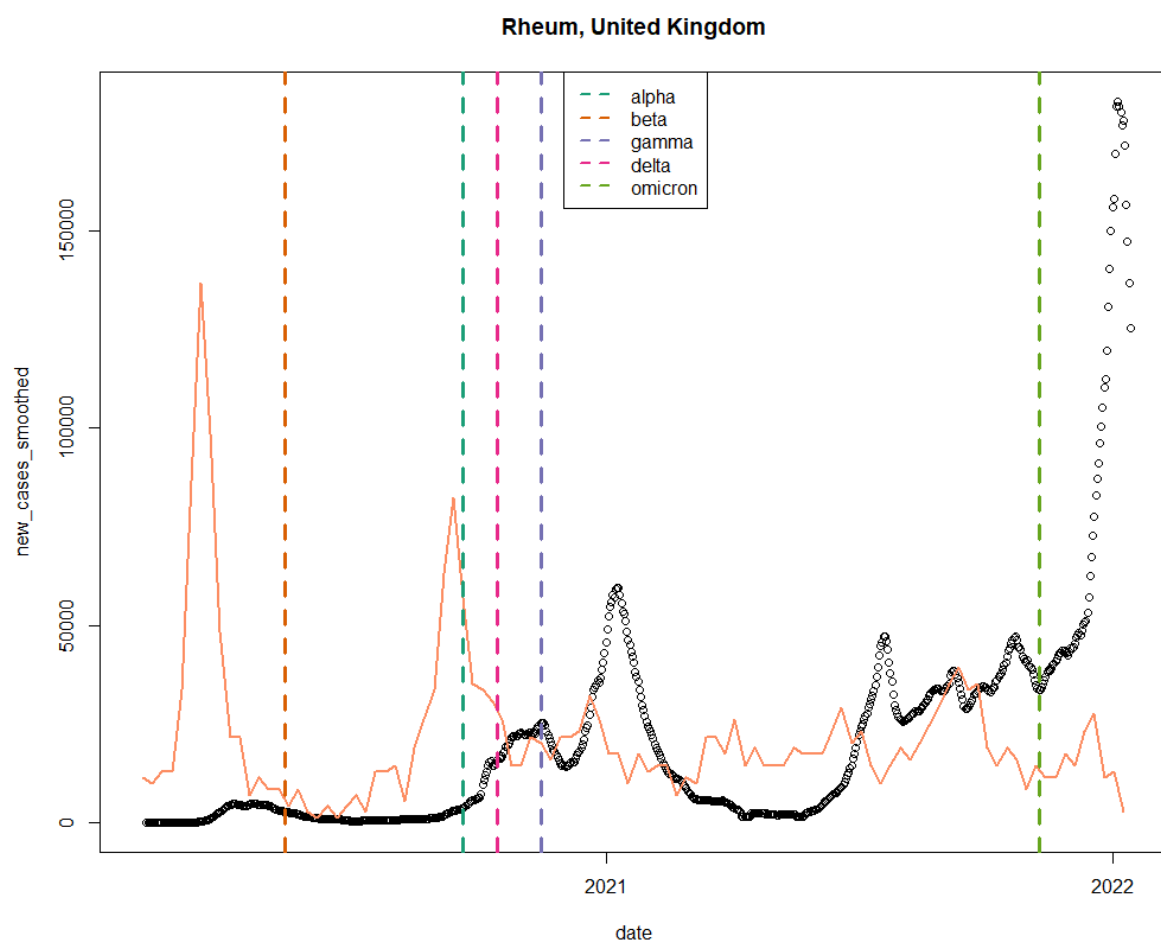


Рисунок 77 Графики популярности запросов по теме Насморк и подтвержденных случаев заболевания в Великобритании

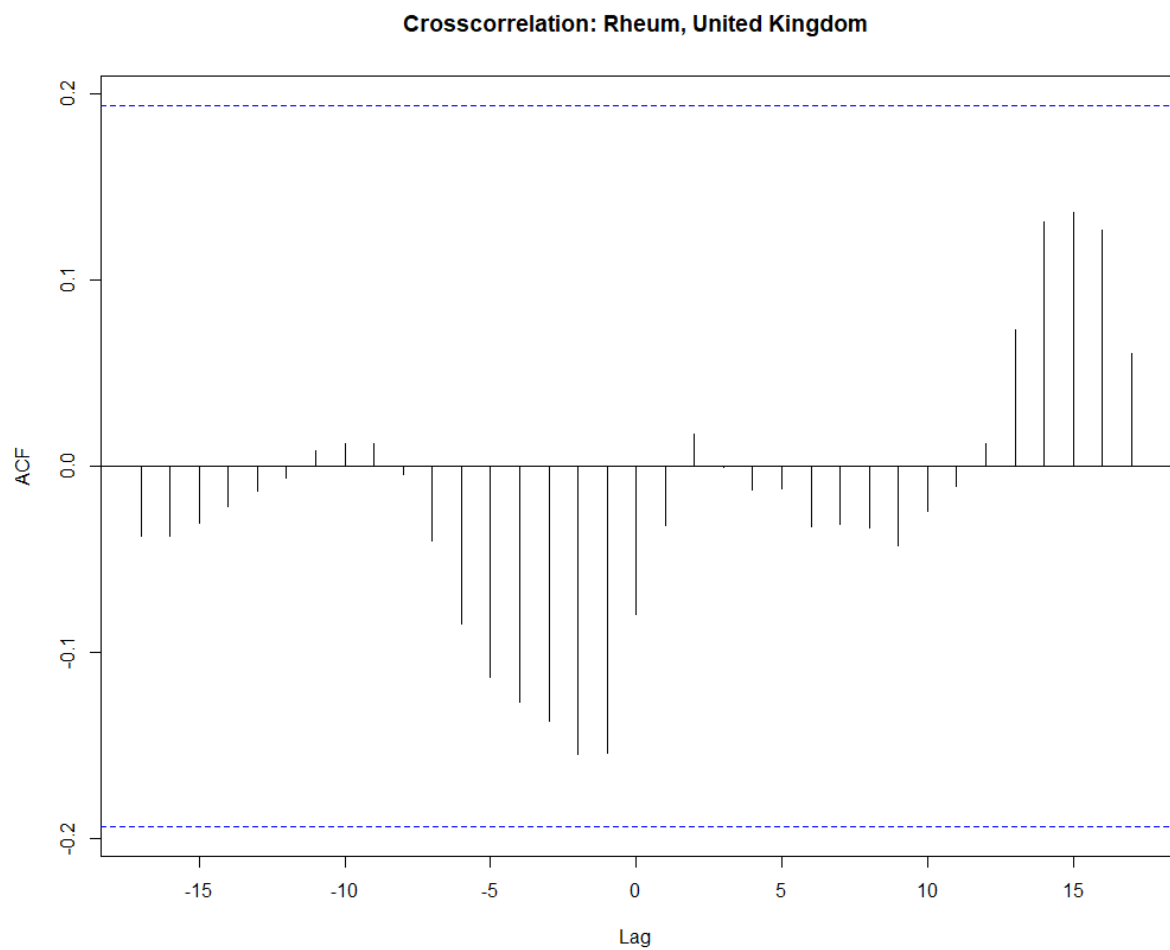


Рисунок 78 График кросскорреляции запросов по теме Насморк и заболеваемостью в Великобритании

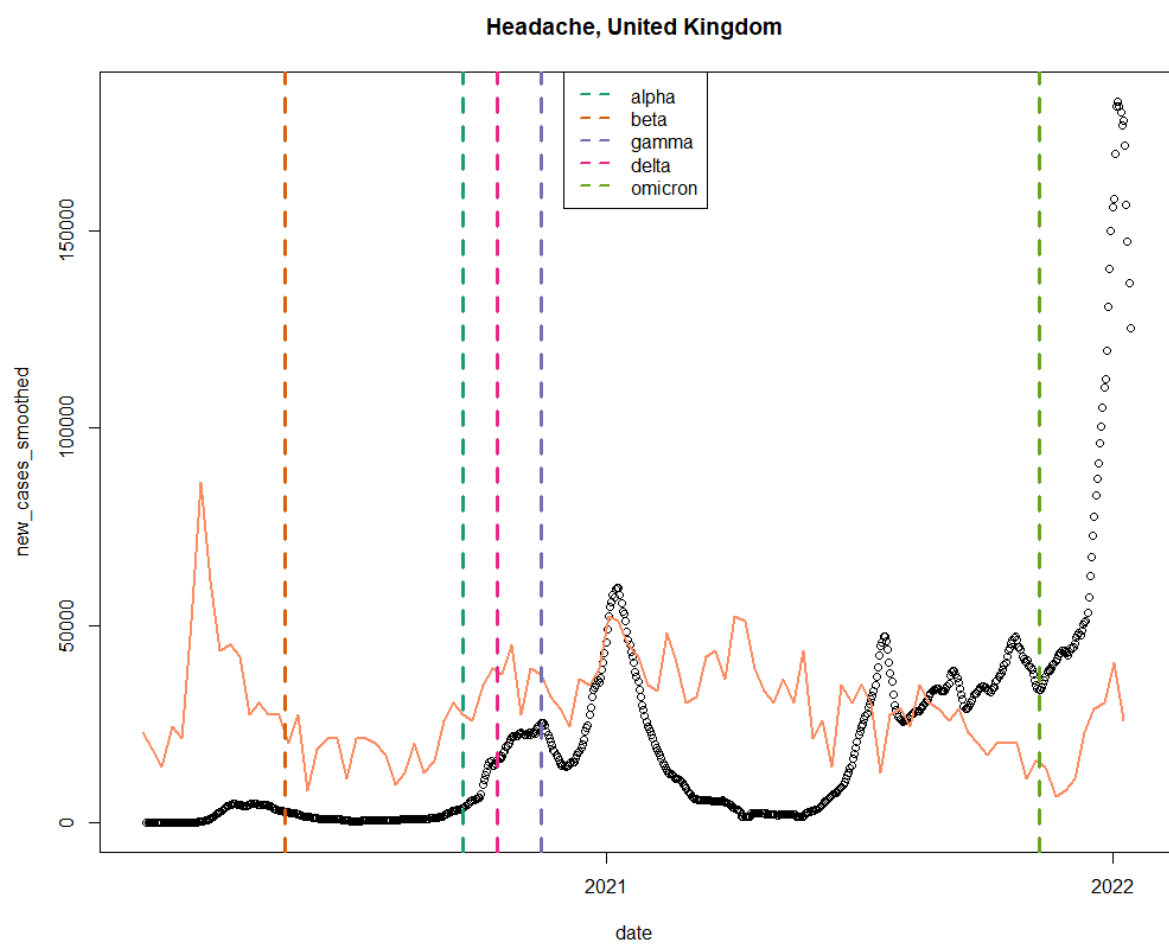


Рисунок 79 Графики популярности запросов по теме Головная боль и подтвержденных случаев заболевания в Великобритании

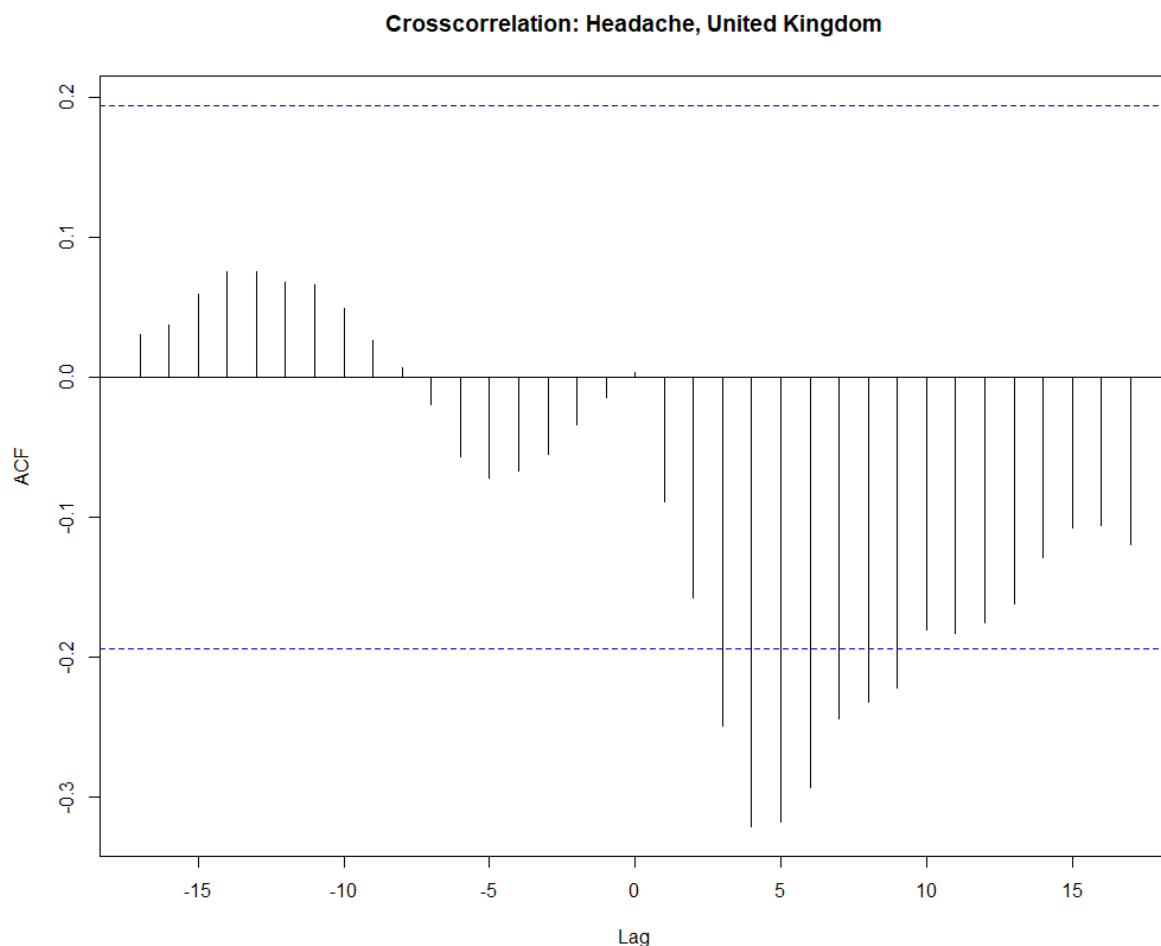


Рисунок 80 График кросскорреляции запросов по теме Головная боль и
заболеваемостью в Великобритании

Анализ по стране

Штамм Альфа впервые зарегистрирован в Великобритании, но ситуация в целом схожа с ситуацией во Франции: так же не зарегистрирована первая волна заражения, очень большой лаг на второй волне.

Примечательно проанализировать последний взрывной рост (который был и во Франции, и в США, и в Индии): рост начинается через некоторое время после первого обнаружения омикрона и можно предположить, что этот рост, как рах обусловлен попаданием нового штамма в страну. Однако если посмотреть на графики популярности запросов, то они почти все (кроме темы «Головная боль») имеют нисходящий тренд во время роста заболеваемости. Возможно, этот штамм легче переносится.

ЮАР

Таблица 7 Корреляция заболеваемости и популярности запросов в ЮАР

WORDKEY	PEARSON	SPEARMAN
COVID	0.592033	0.546078
SMELL	0.571574	0.541808
THROAT	0.421824	0.448668
FEWER	0.311262	0.234764
DYSPNEA	0.311348	0.269788
COUGH	0.568259	0.400261
RHEUM	0.155989	0.263895
HEADACHE	0.379772	0.272044

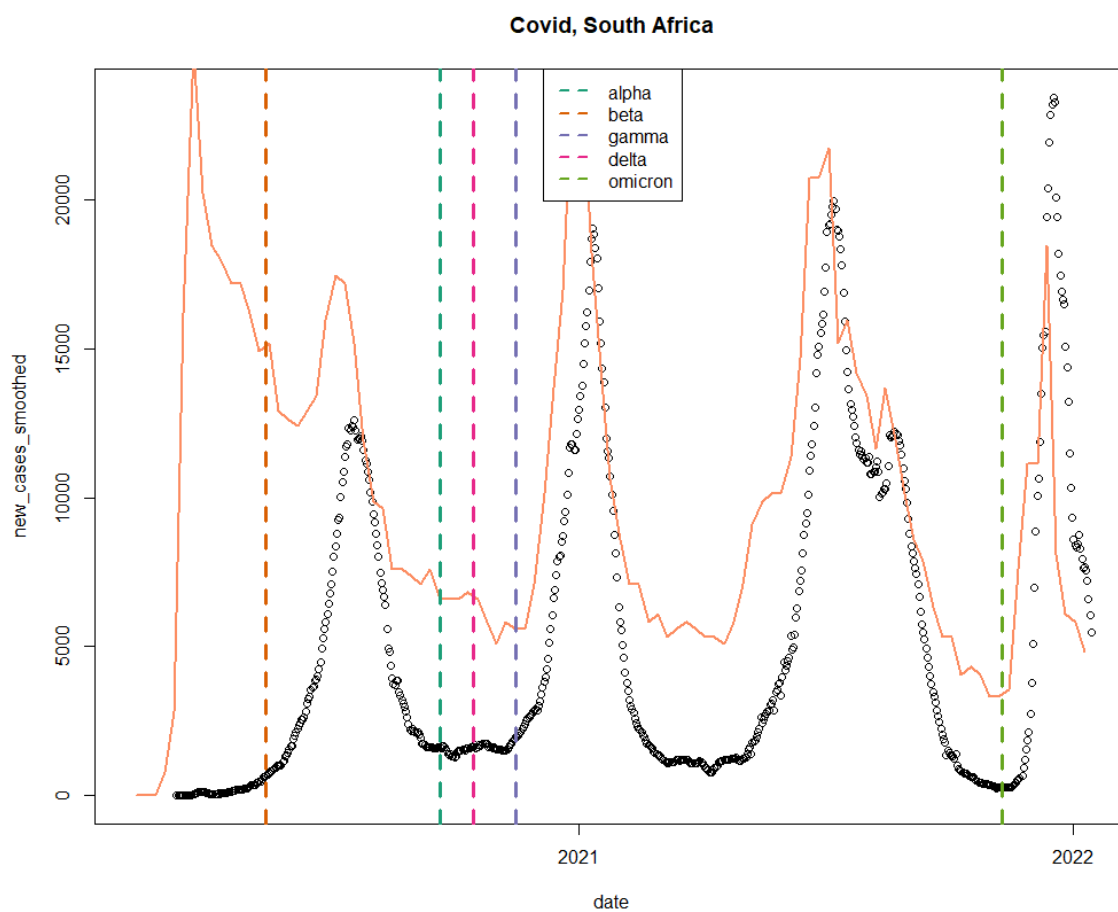


Рисунок 81 Графики популярности запросов по теме COVID-19 и подтвержденных случаев заболевания в ЮАР

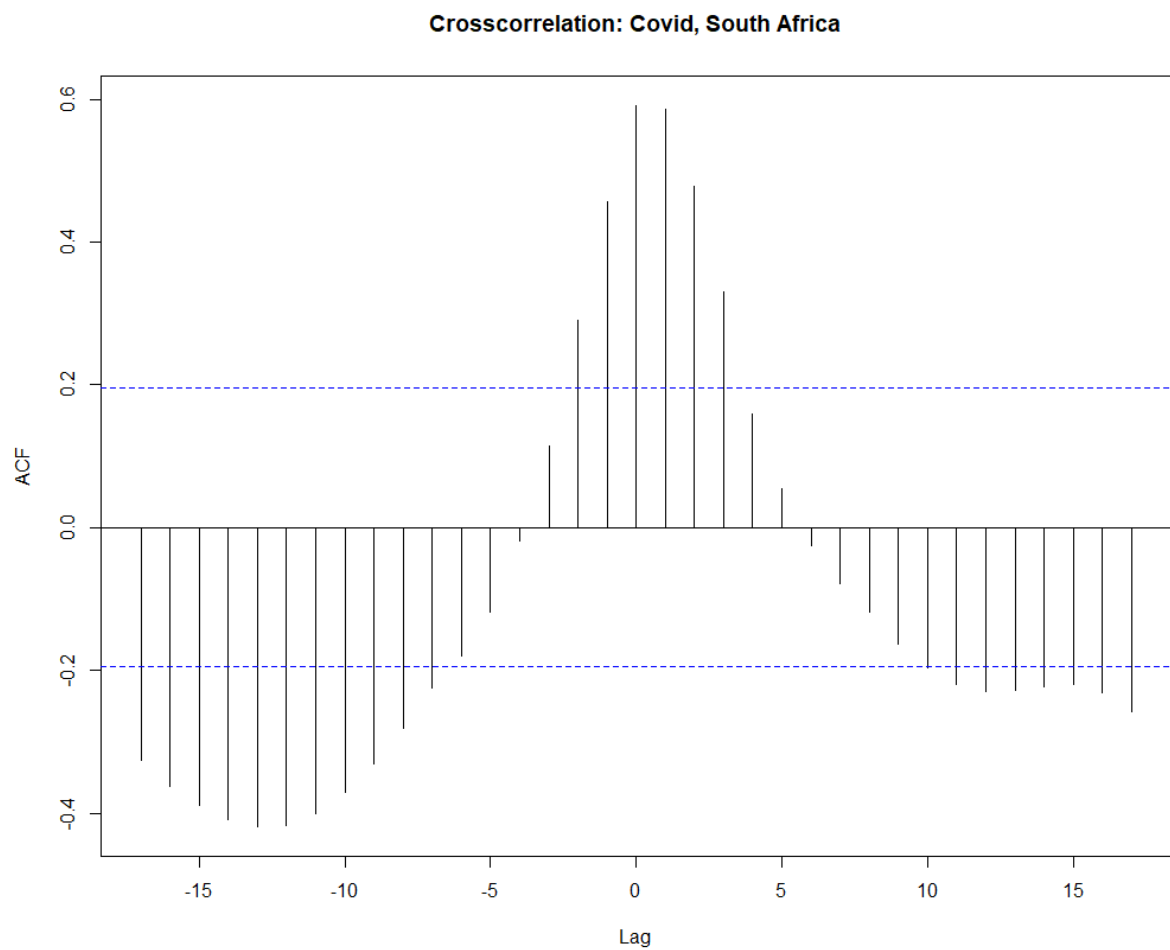


Рисунок 82 График кросскорреляции запросов по теме COVID-19 и заболеваемостью в ЮАР

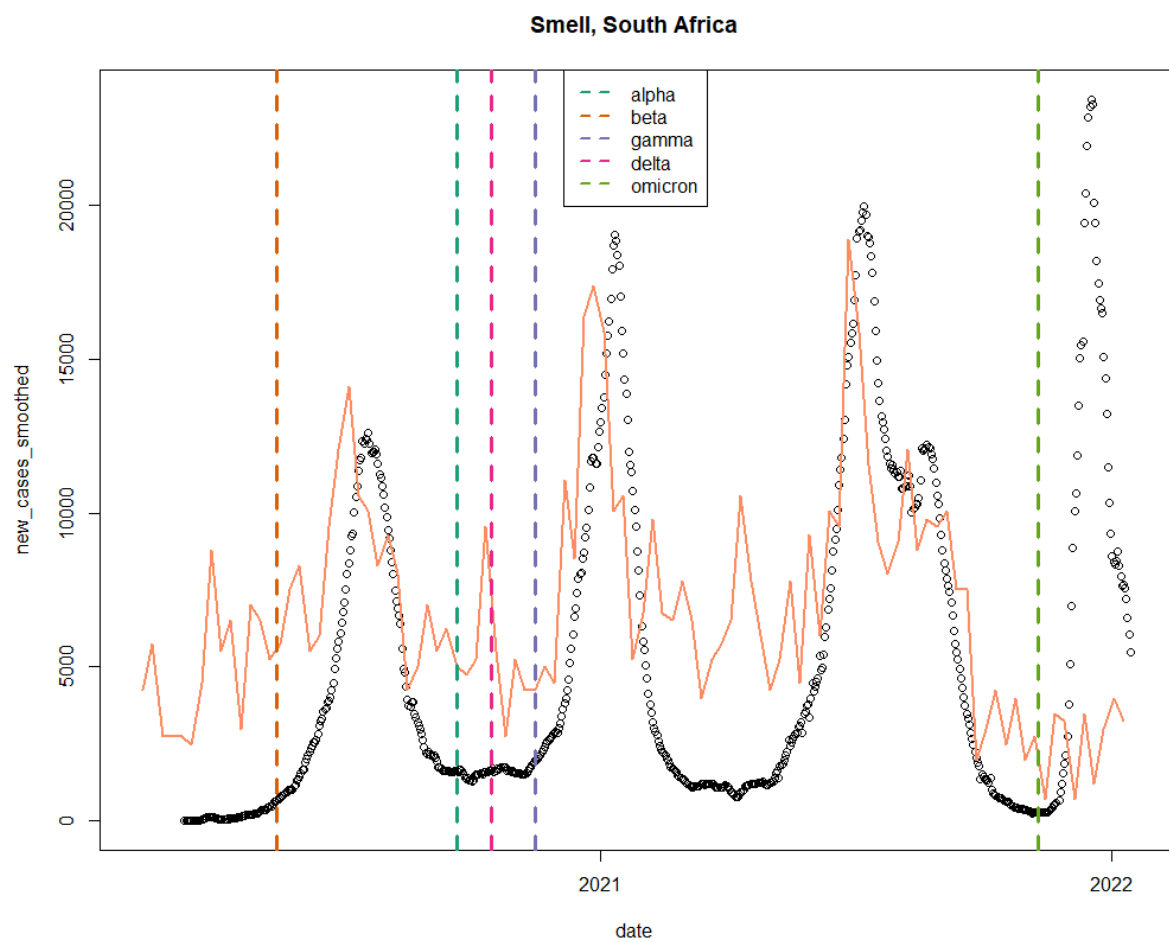


Рисунок 83 Графики популярности запросов по теме Обоняние и подтвержденных случаев заболевания в ЮАР

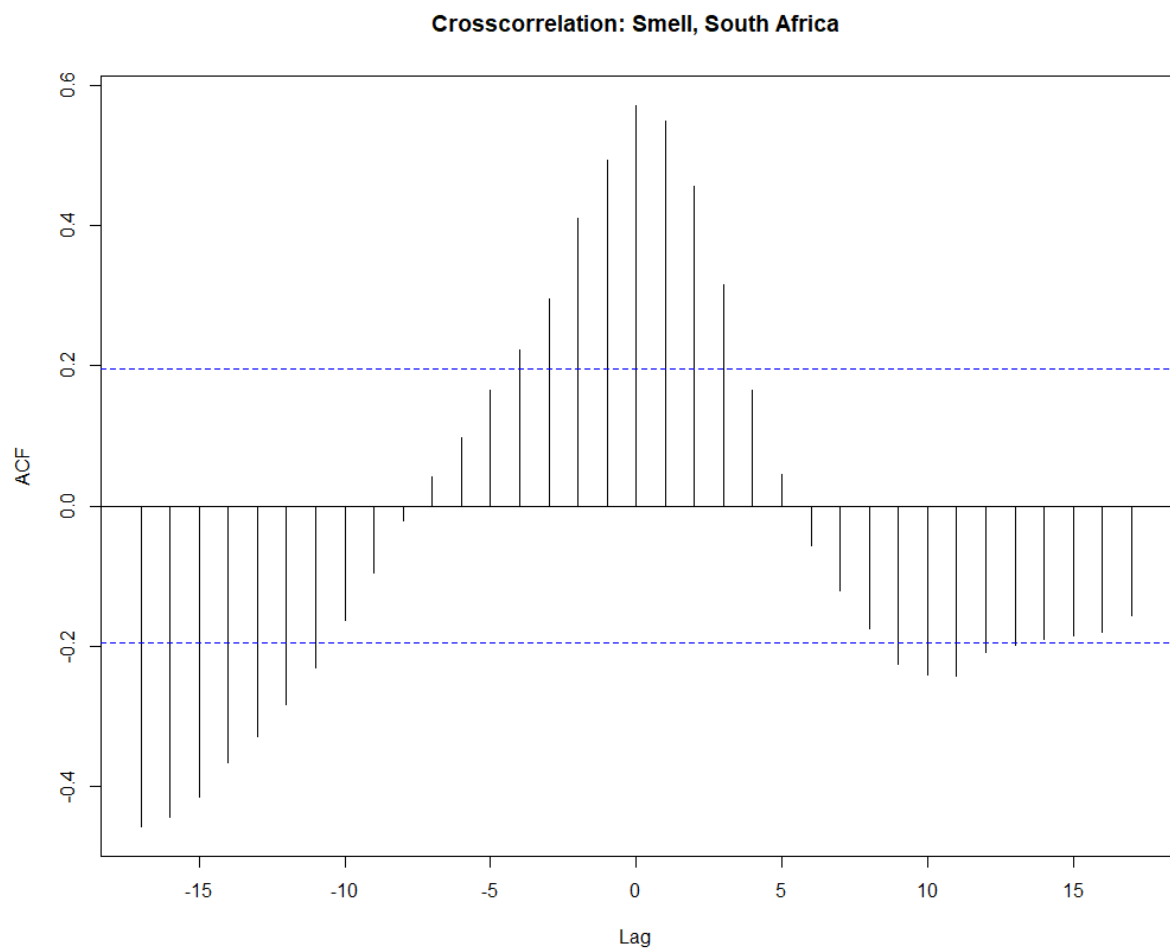


Рисунок 84 График кросскорреляции запросов по теме Обоняние и заболеваемостью в ЮАР

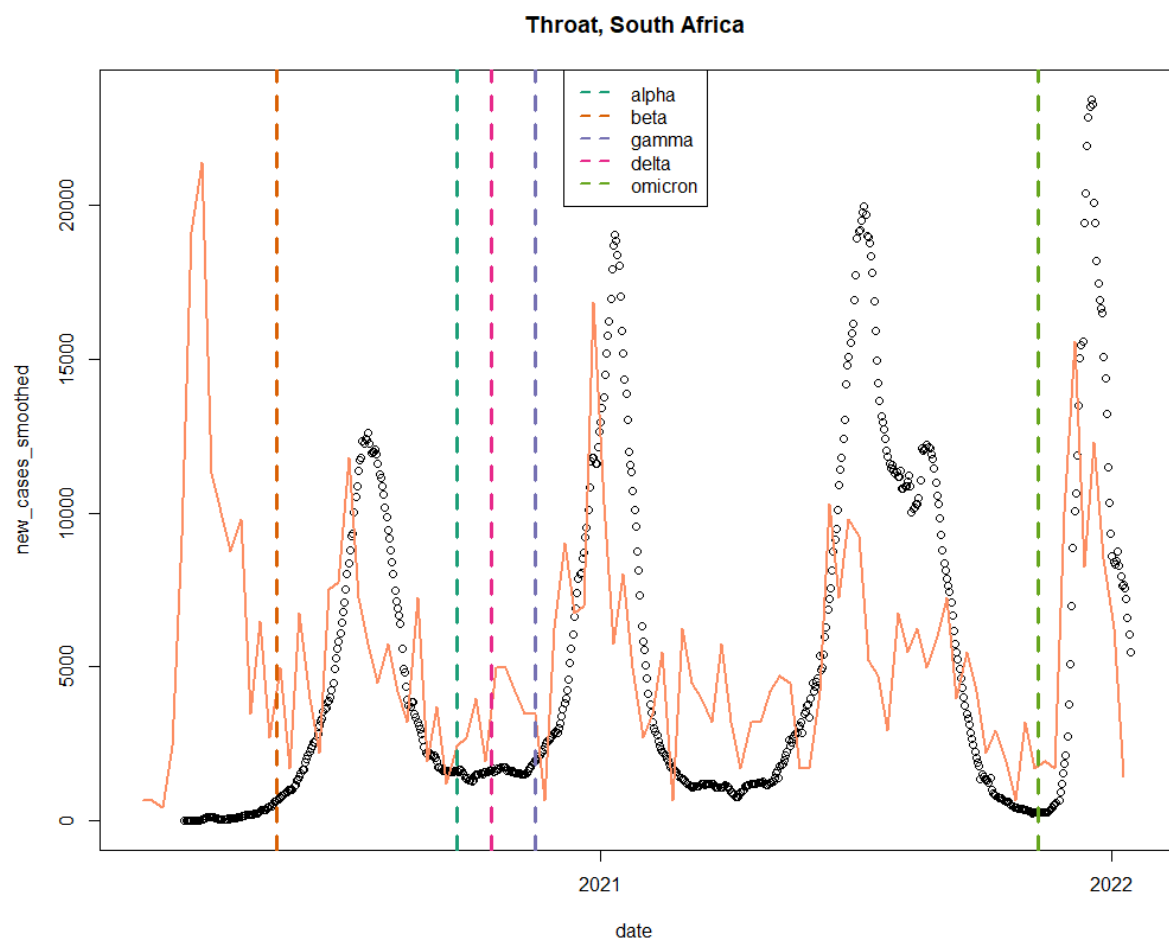


Рисунок 85 Графики популярности запросов по теме Больное горло и подтвержденных случаев заболевания в ЮАР

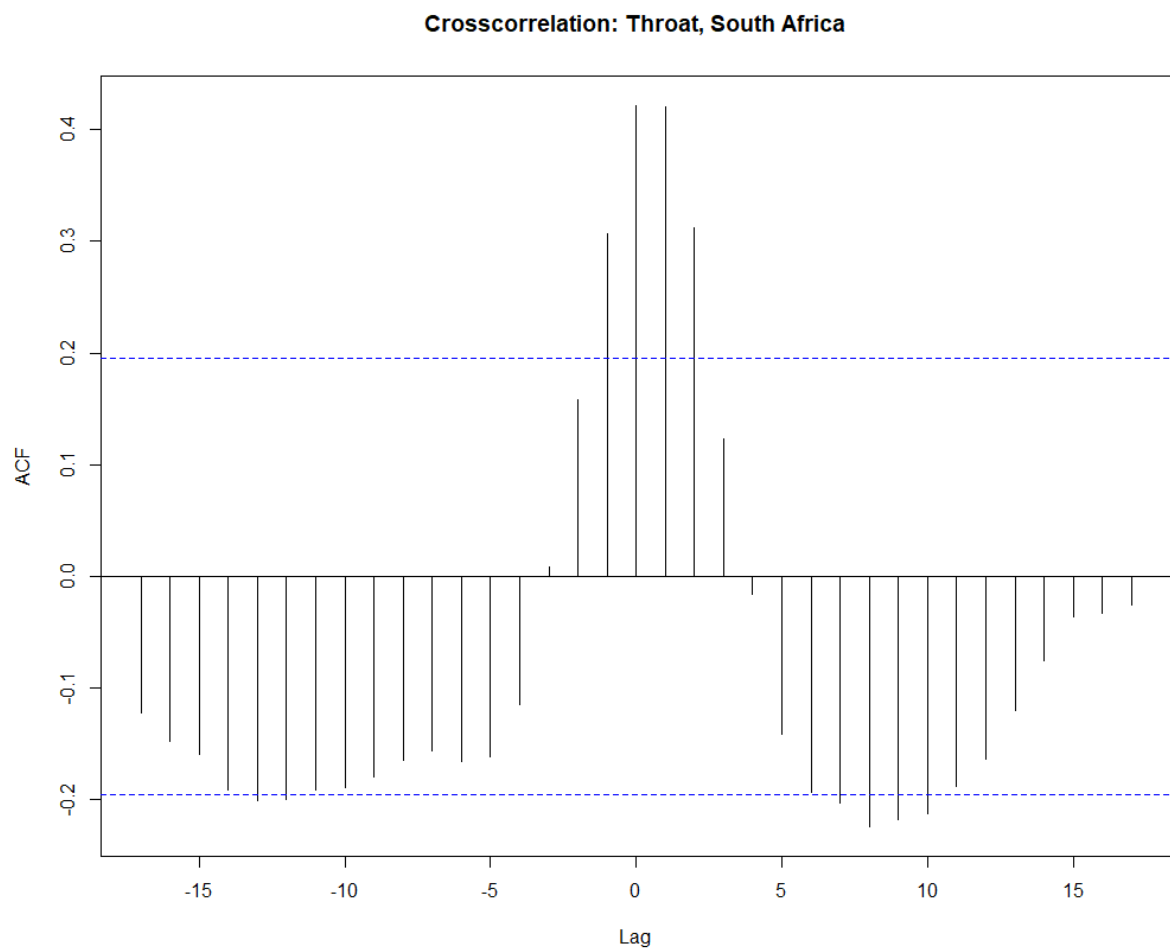


Рисунок 86 График кросскорреляции запросов по теме Больное горло и
заболеваемостью в ЮАР

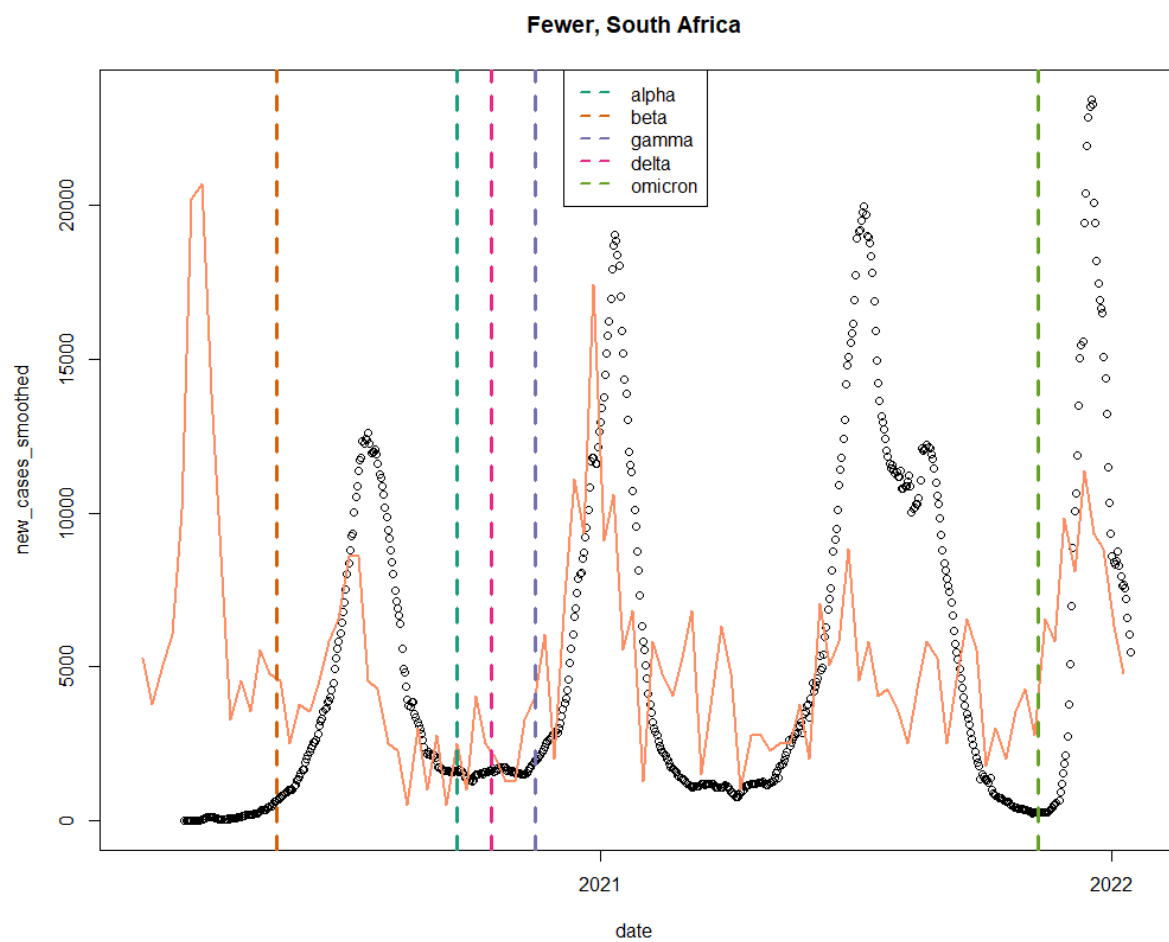


Рисунок 87 Графики популярности запросов по теме Лихорадка и подтвержденных случаев заболевания в ЮАР

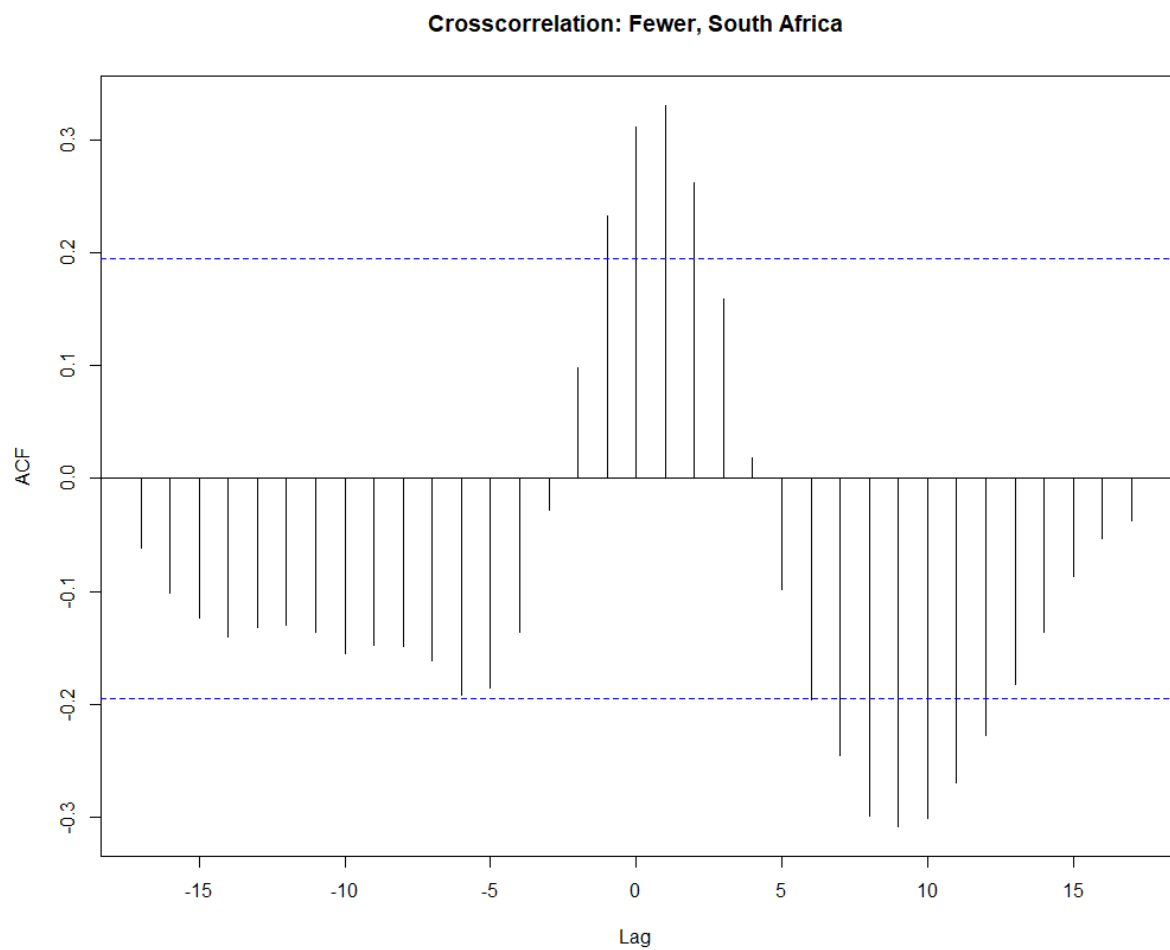


Рисунок 88 График кросскорреляции запросов по теме Лихорадка и заболеваемостью в ЮАР

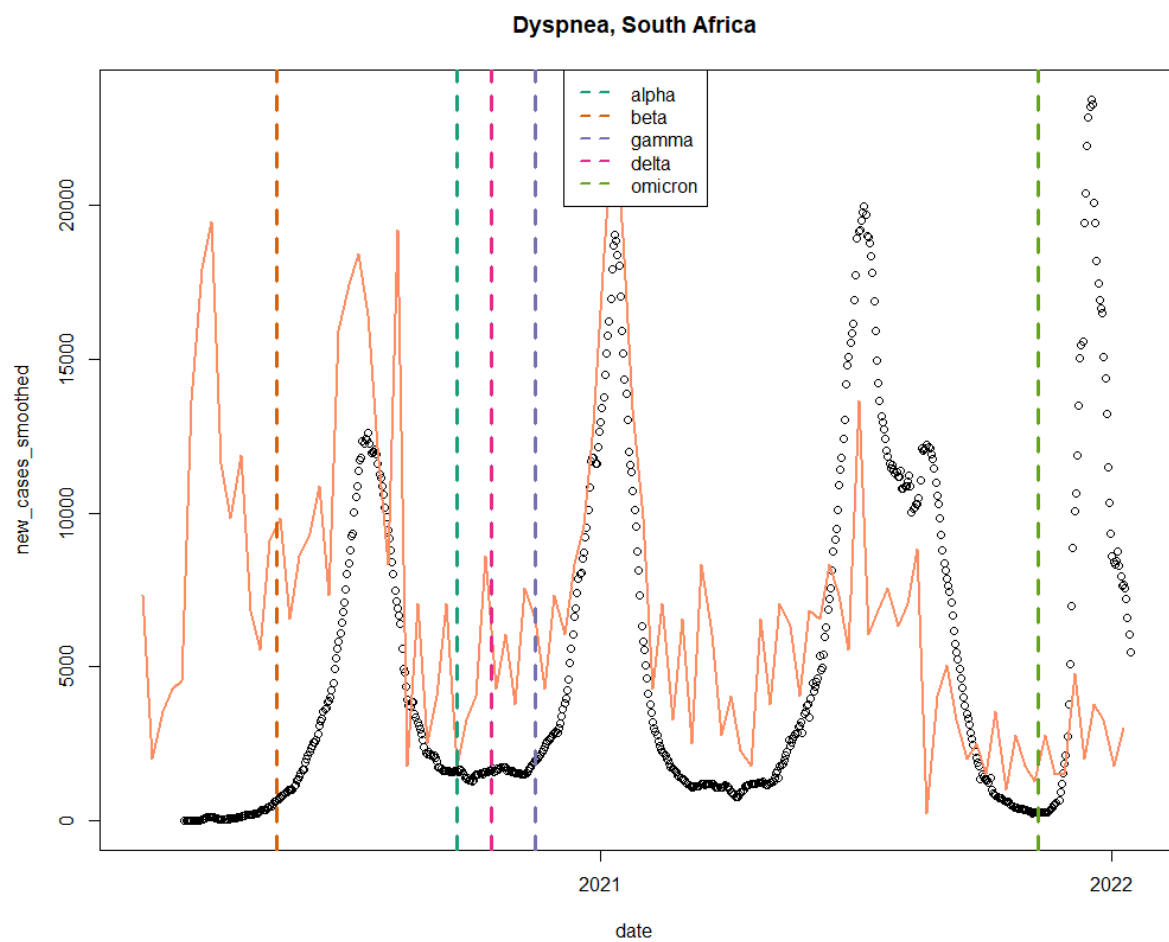


Рисунок 89 Графики популярности запросов по теме Обоняние и подтвержденных случаев заболевания в ЮАР

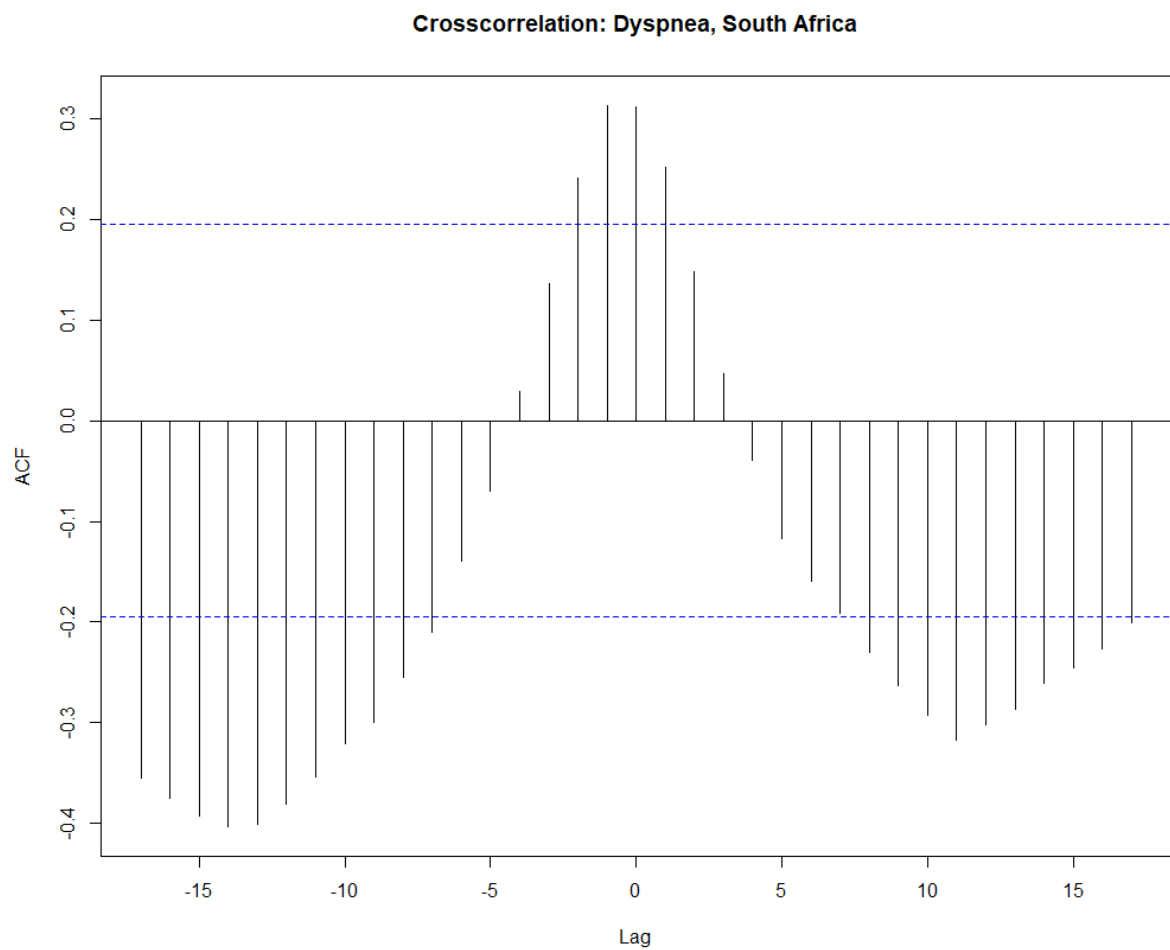


Рисунок 90 График кросскорреляции запросов по теме Одышка и заболеваемостью в ЮАР

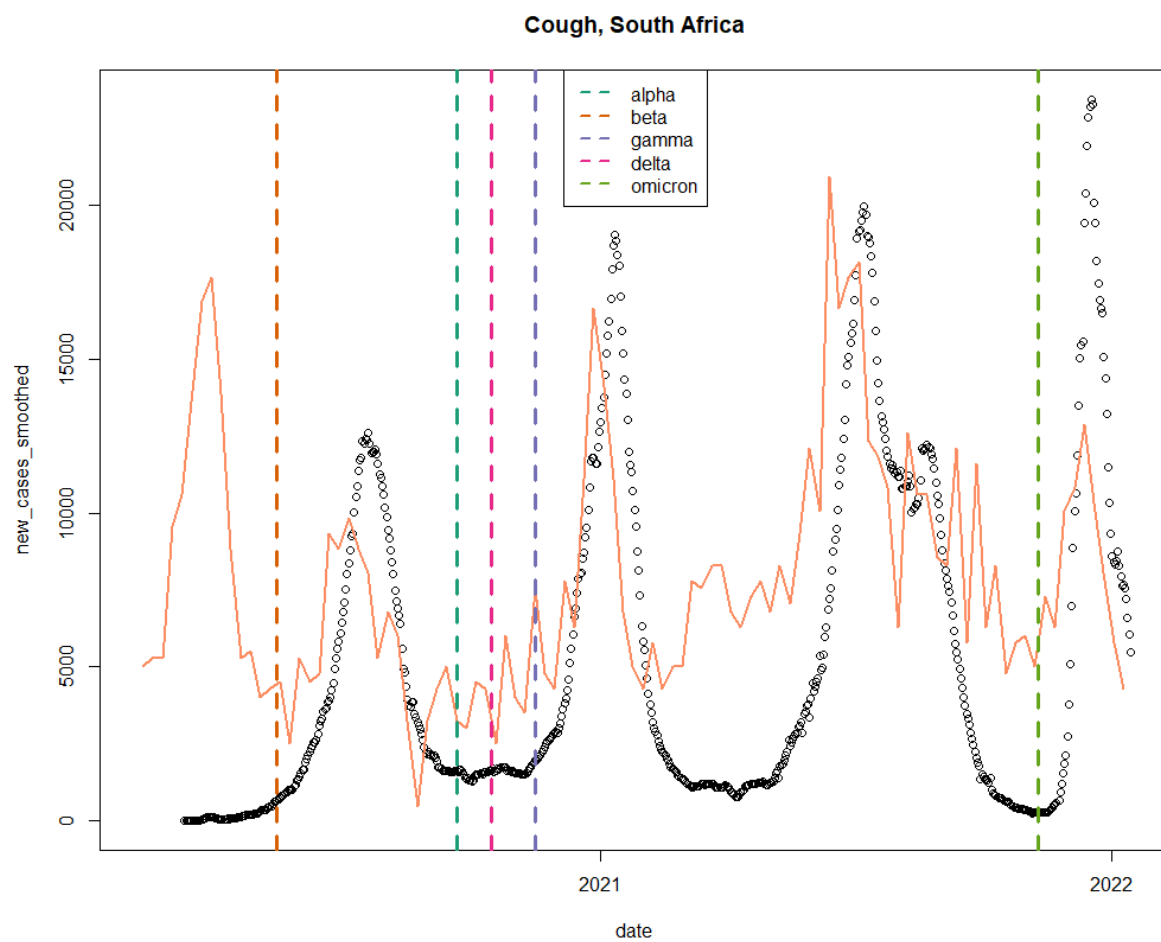


Рисунок 91 Графики популярности запросов по теме Кашель и подтвержденных случаев заболевания в ЮАР

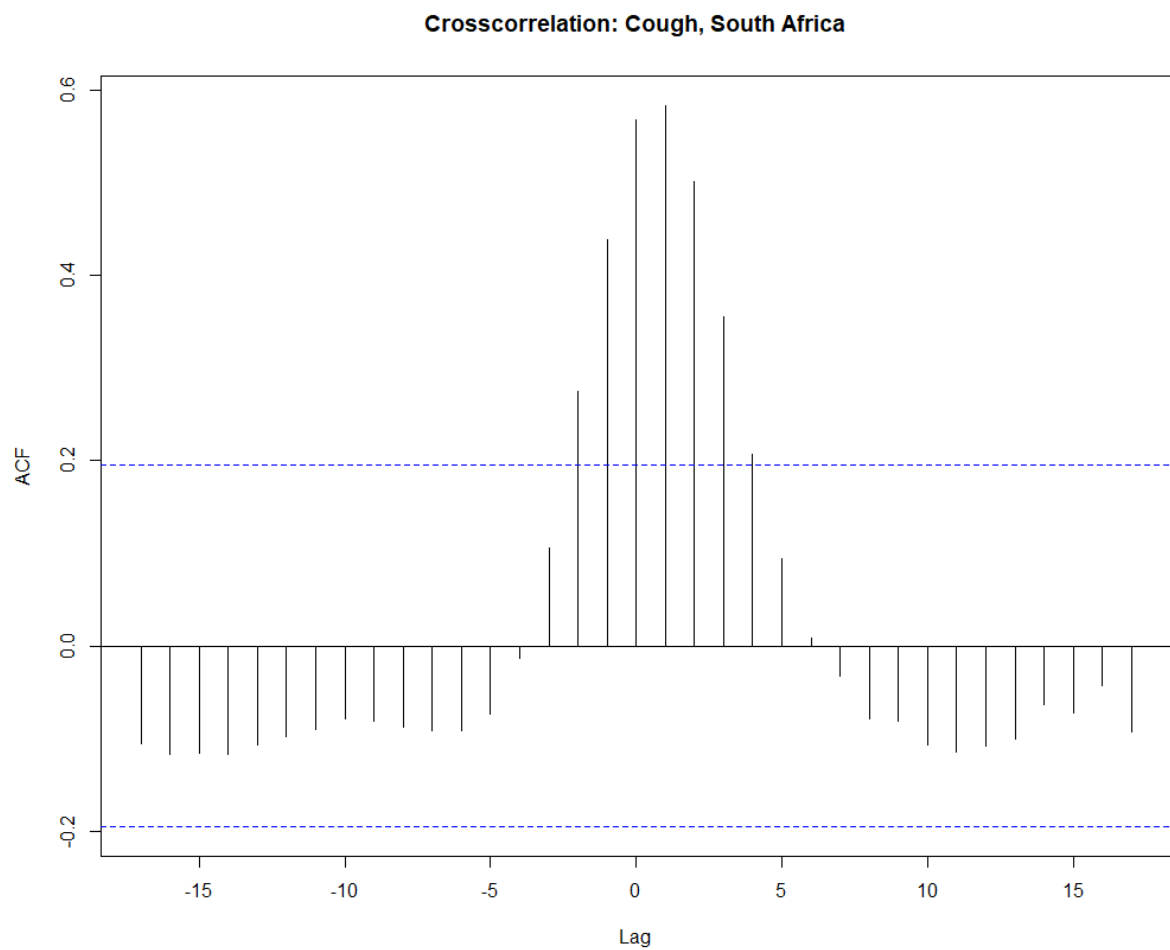


Рисунок 92 График кросскорреляции запросов по теме Кашель и заболеваемостью в ЮАР

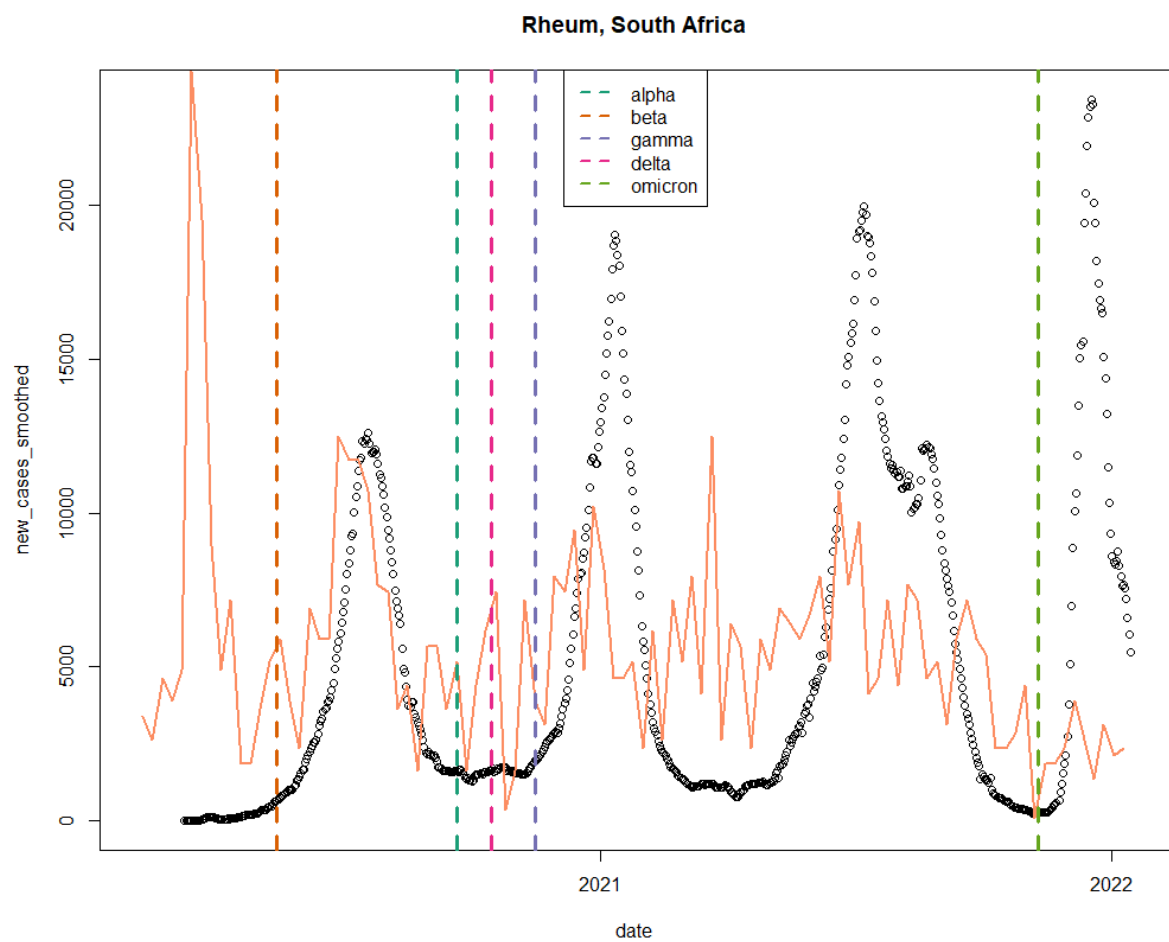


Рисунок 93 Графики популярности запросов по теме Насморк и подтвержденных случаев заболевания в ЮАР

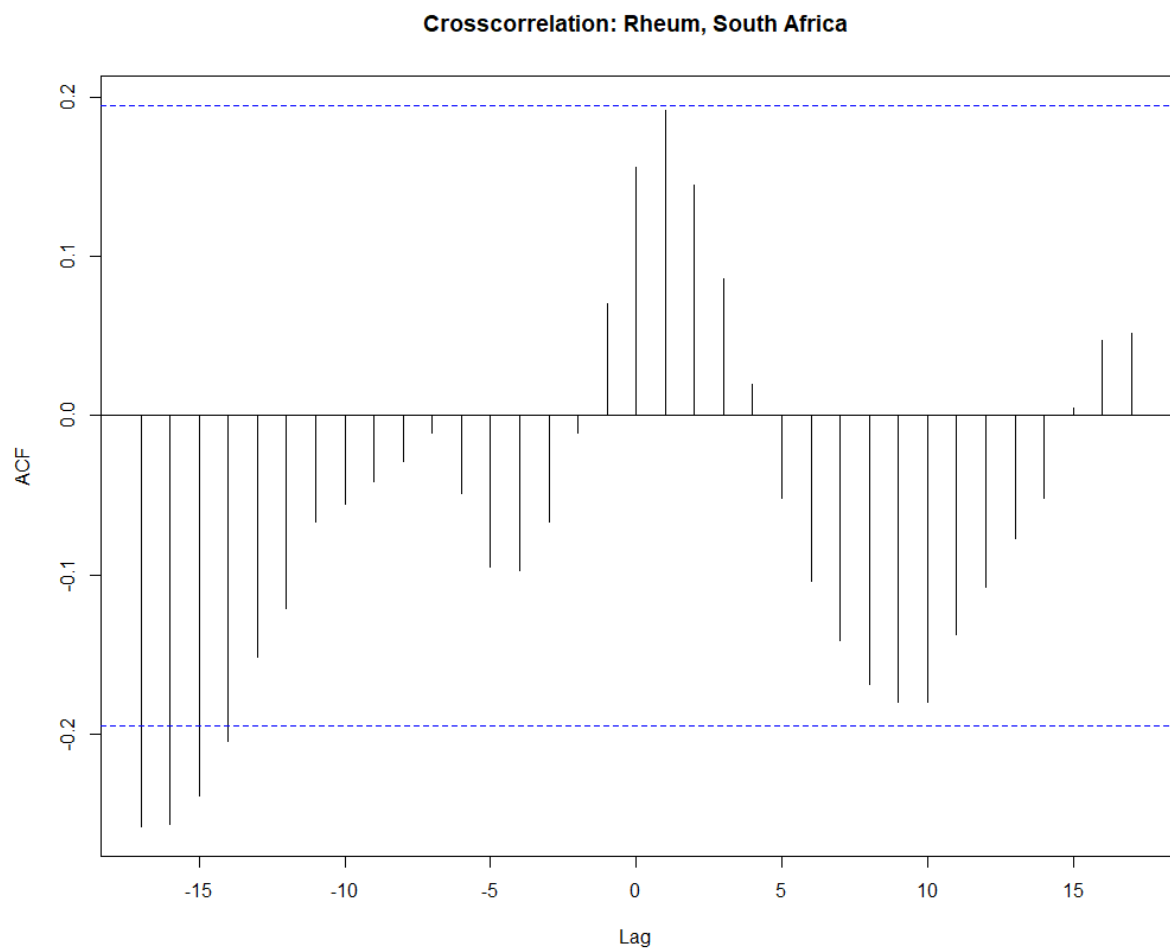


Рисунок 94 График кросскорреляции запросов по теме Насморк и заболеваемостью в
ЮАР

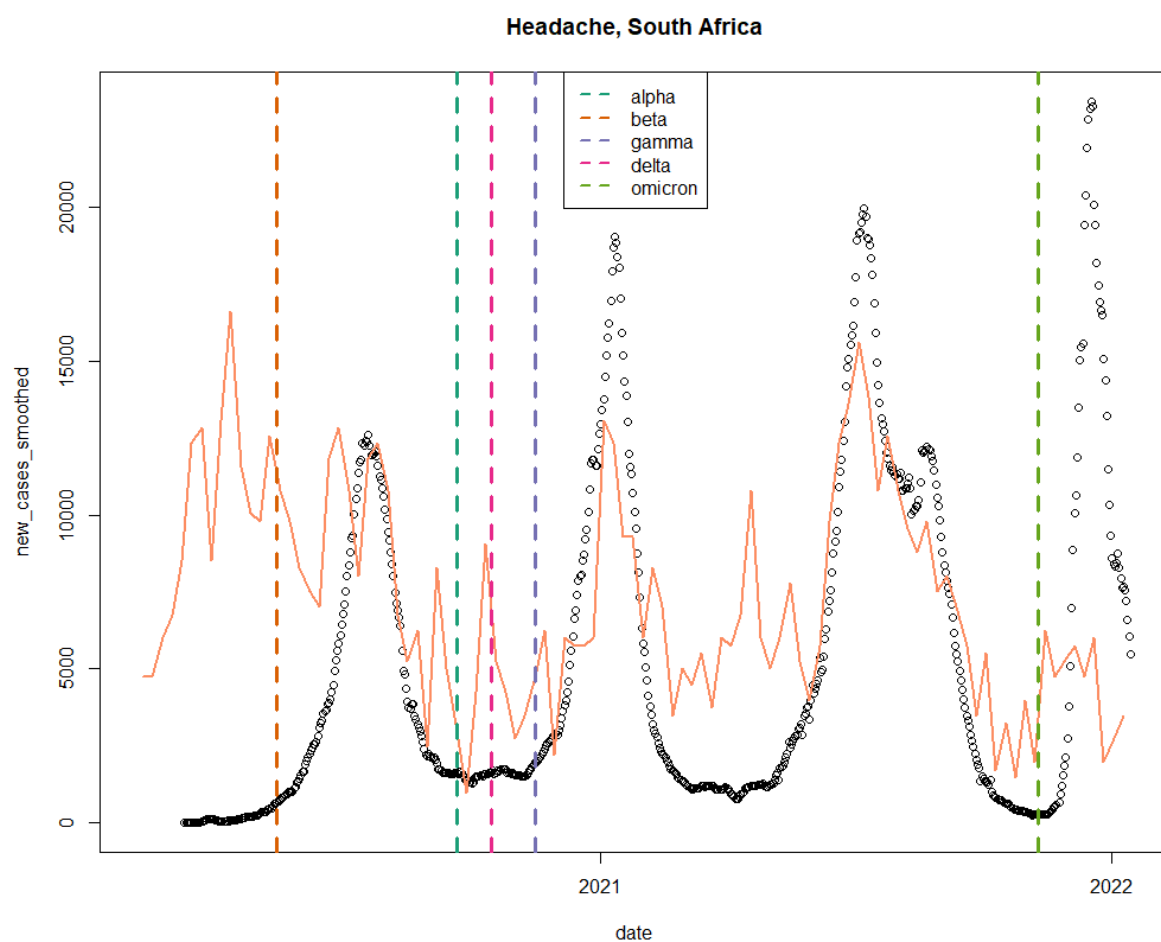


Рисунок 95 Графики популярности запросов по теме Головная боль и подтвержденных случаев заболевания в ЮАР

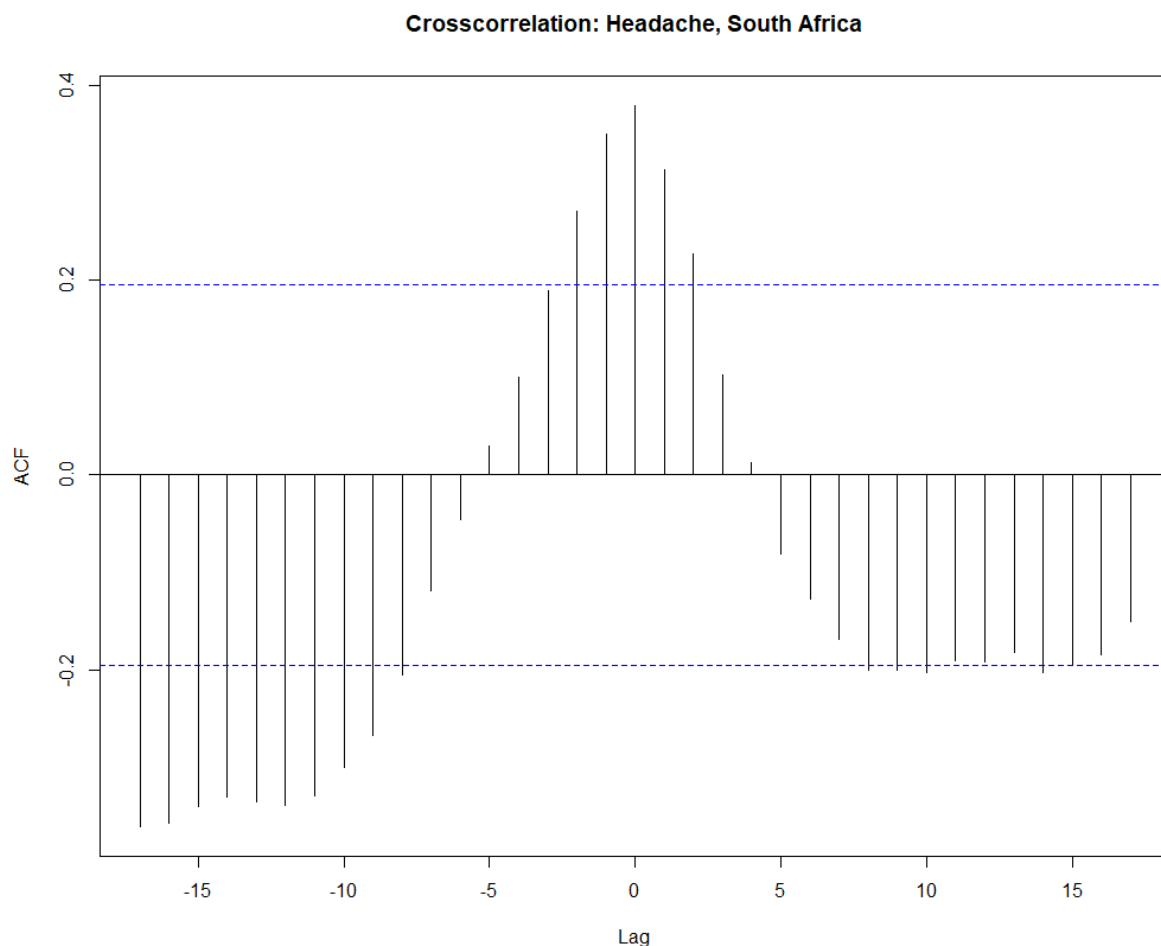


Рисунок 96 График кросскорреляции запросов по теме Головная боль и заболеваемостью в ЮАР

Анализ по стране:

В ЮАРе зарегистрированы первые случаи заражения штаммами бета и омикрон. Как видно из графиков, ЮАР уже завершает волну заражений омикроном. В ЮАРе, как и в Индии очень высокая корреляция числа заражений и популярностей запросов.

Насчет омикрона можно предположить, что обоняние при нем не теряется, болит горло, поднимается температура, нет одышки, есть кашель, нет насморка.

Бразилия

Таблица 8 Корреляция заболеваемости и популярности запросов во Франции

WORDKEY	PEARSON	SPEARMAN
COVID	-0.15773	0.275985
SMELL	0.298093	0.369626
THROAT	-0.21112	-0.06871
FEWER	-0.17313	0.022171
DYSPNEA	-0.03097	0.293709
COUGH	-0.29896	-0.2442
RHEUM	0.111077	0.238624
HEADACHE	0.437231	0.384068

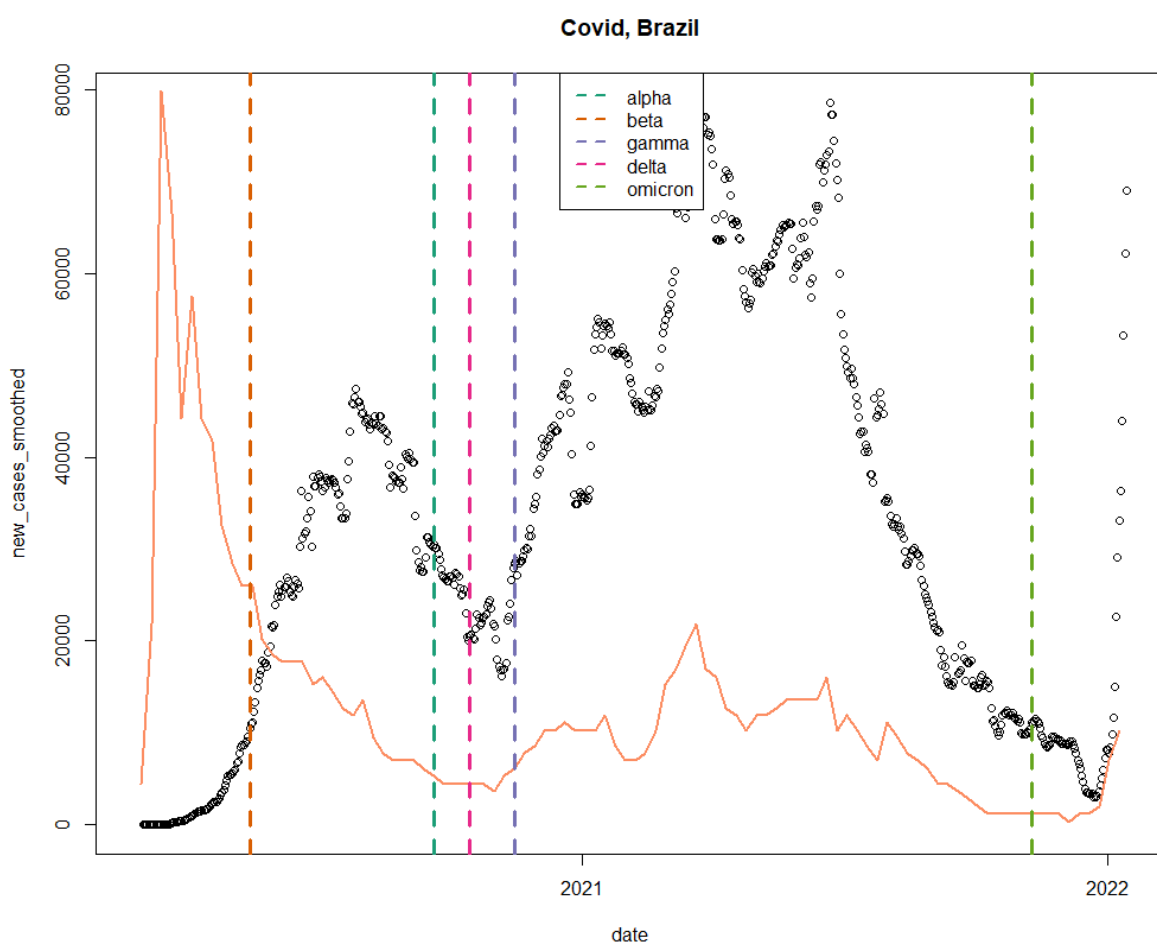


Рисунок 97 Графики популярности запросов по теме COVID-19 и подтвержденных случаев заболевания в Бразилии

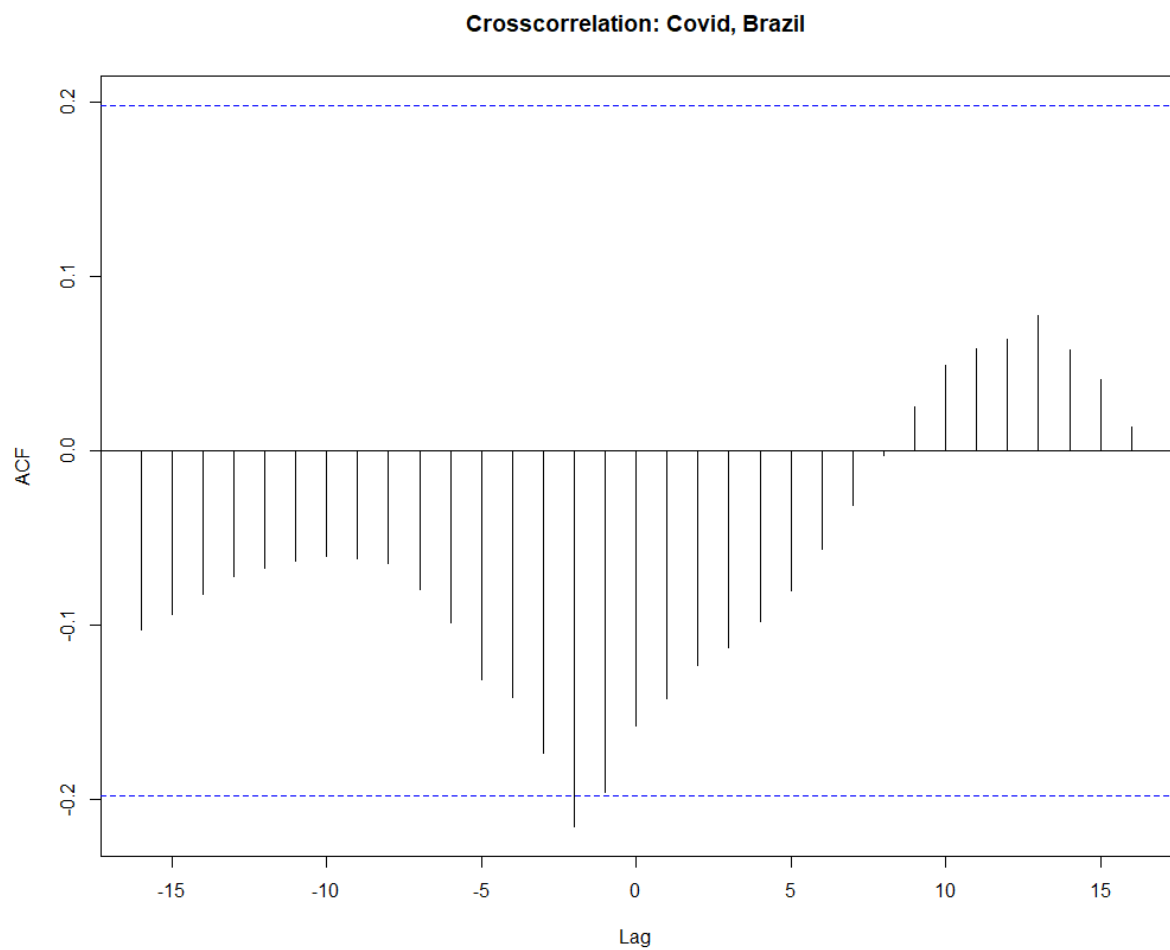


Рисунок 98 График кросскорреляции запросов по теме COVID-19 и заболеваемостью в Бразилии

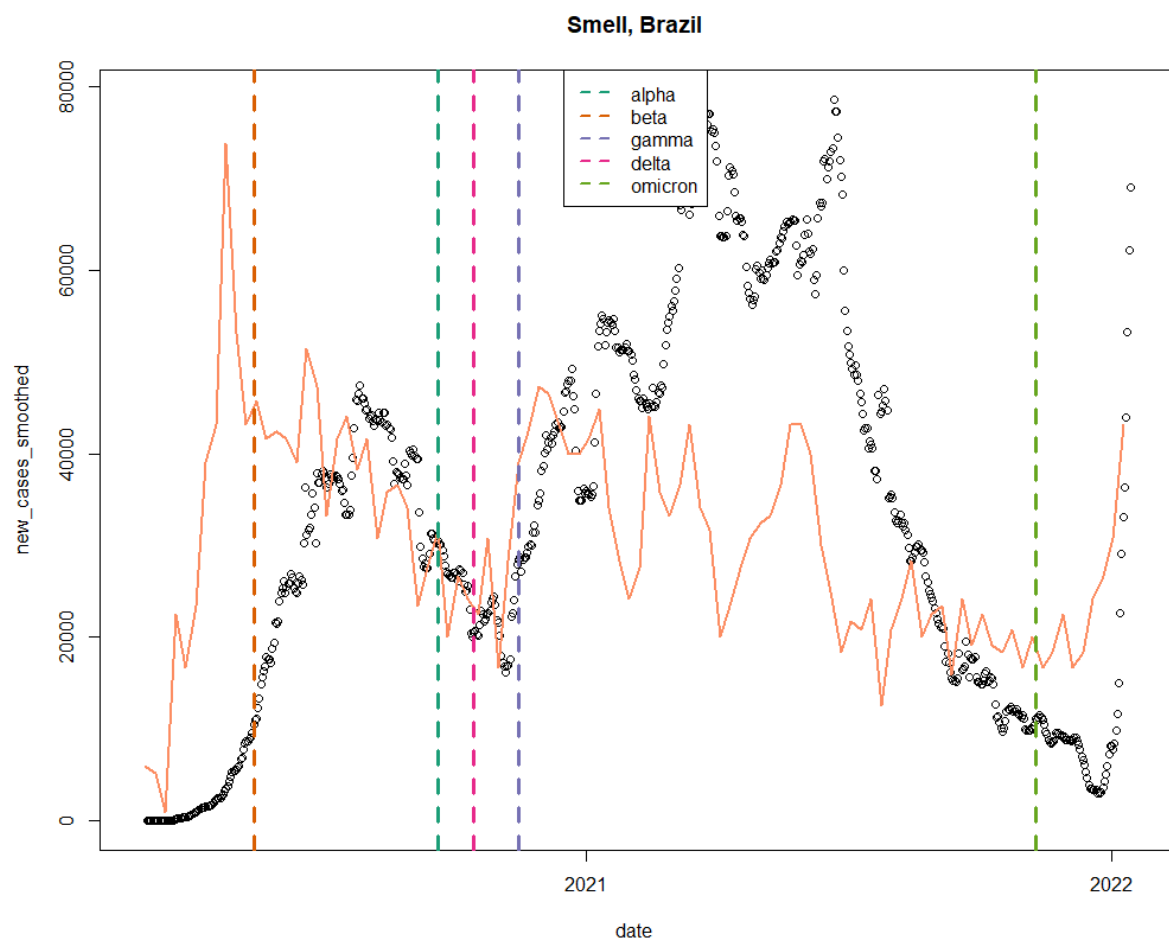


Рисунок 99 Графики популярности запросов по теме Обоняние и подтвержденных случаев заболевания в Бразилии

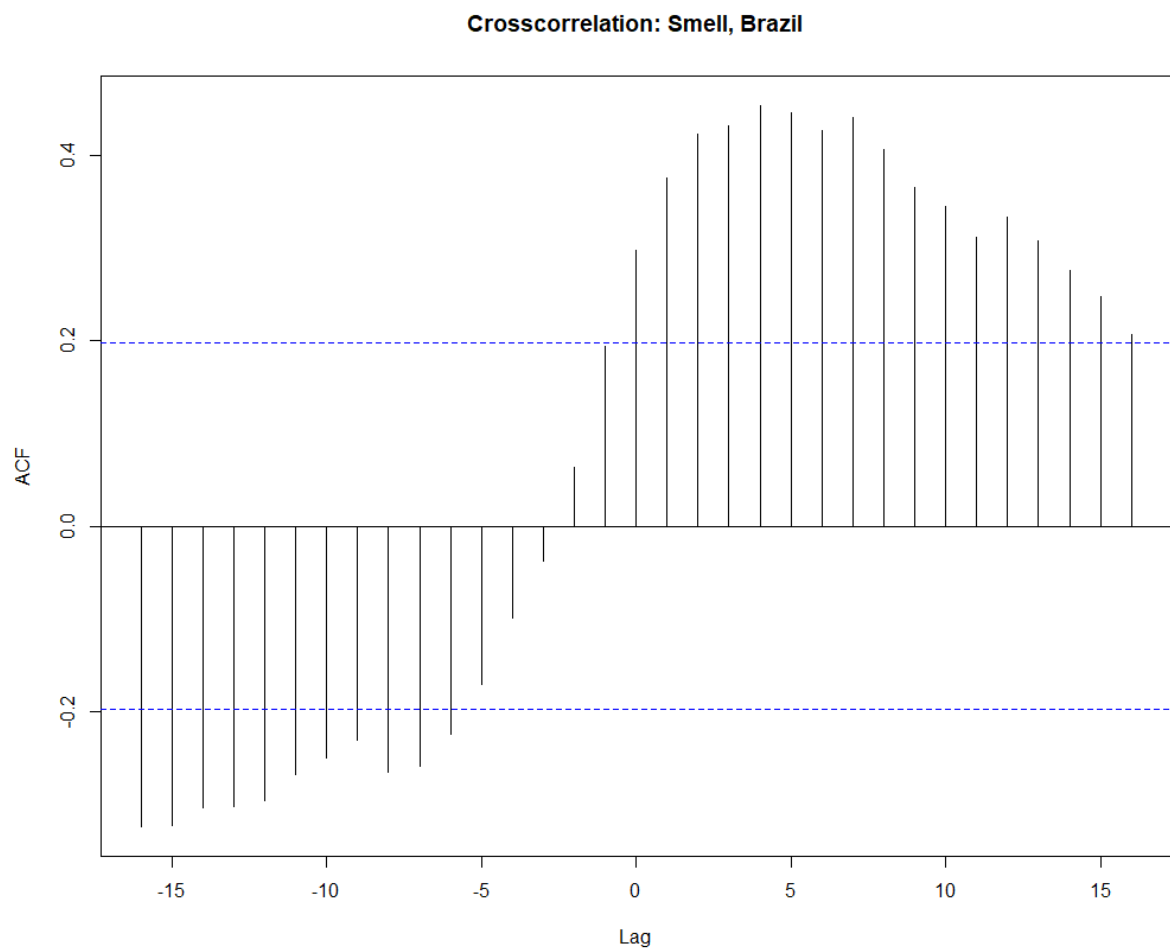


Рисунок 100 График кросскорреляции запросов по теме Обоняние и заболеваемостью в Бразилии

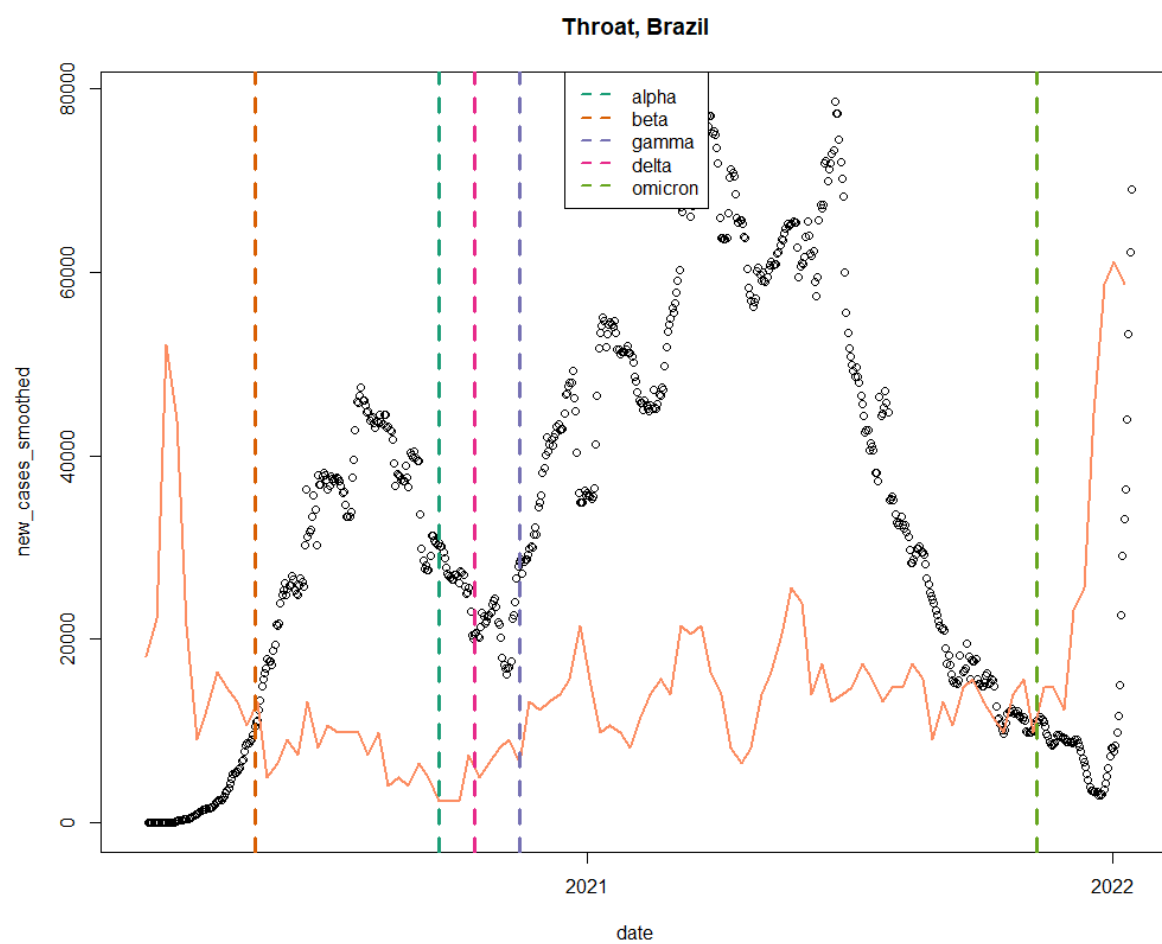


Рисунок 101 Графики популярности запросов по теме Больное горло и подтвержденных случаев заболевания в Бразилии

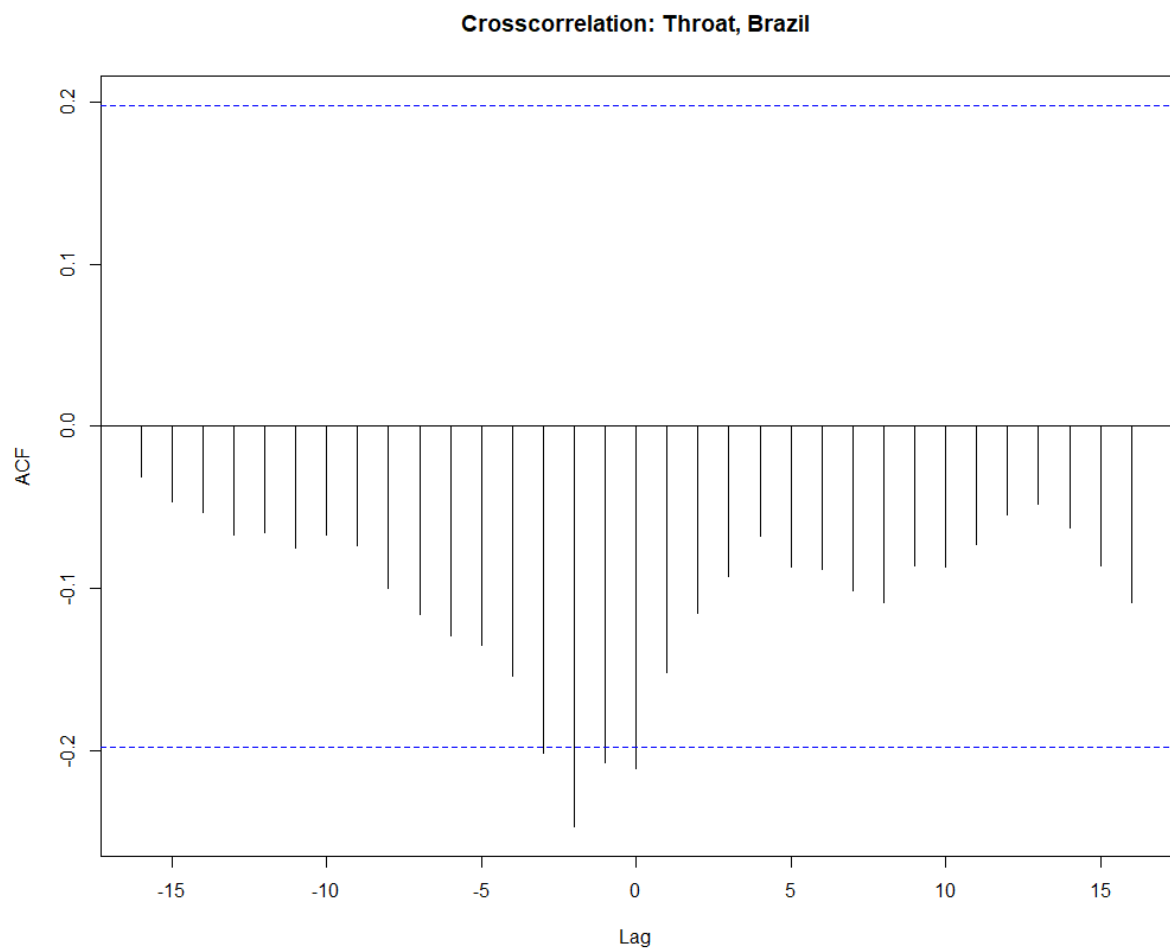


Рисунок 102 График кросскорреляции запросов по теме Больное горло и заболеваемостью в Бразилии

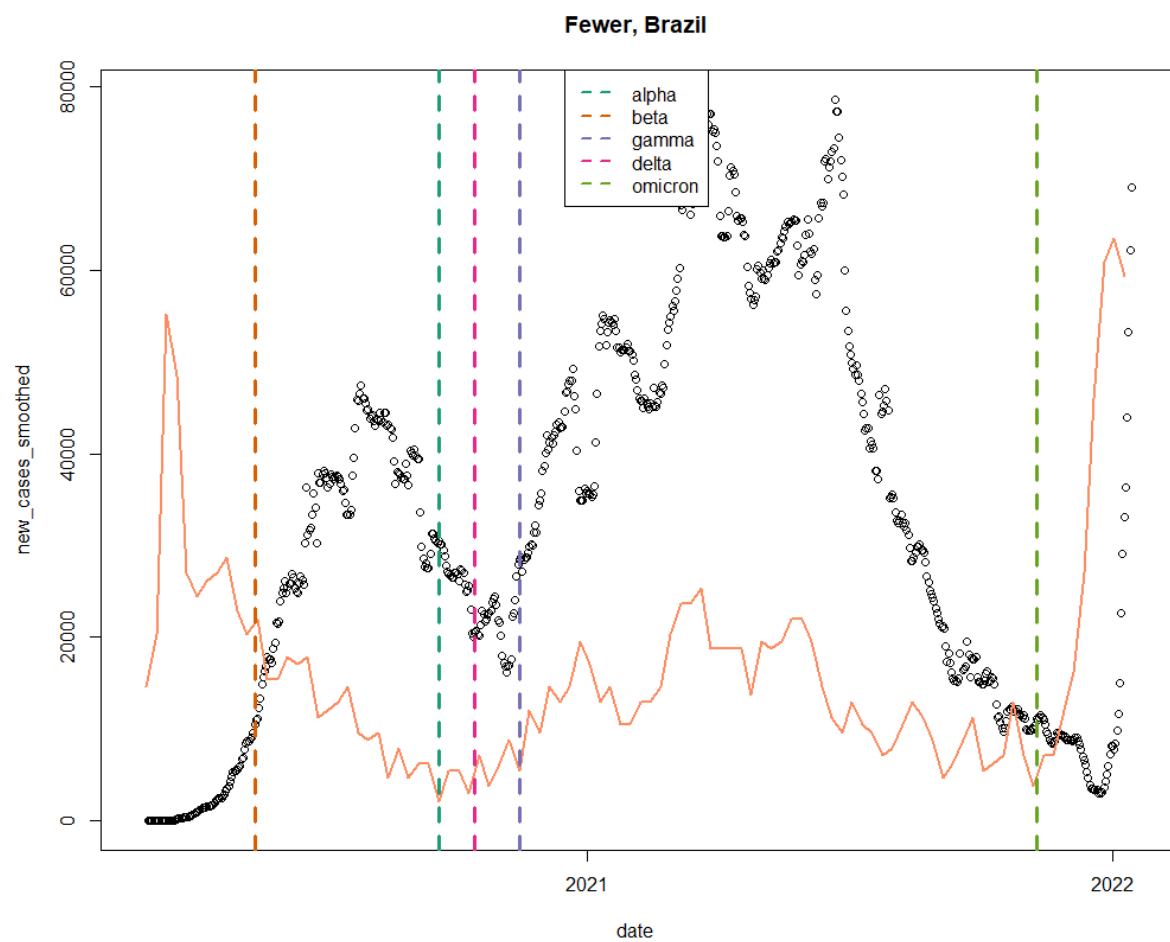


Рисунок 103 Графики популярности запросов по теме Лихорадка и подтвержденных случаев заболевания в Бразилии

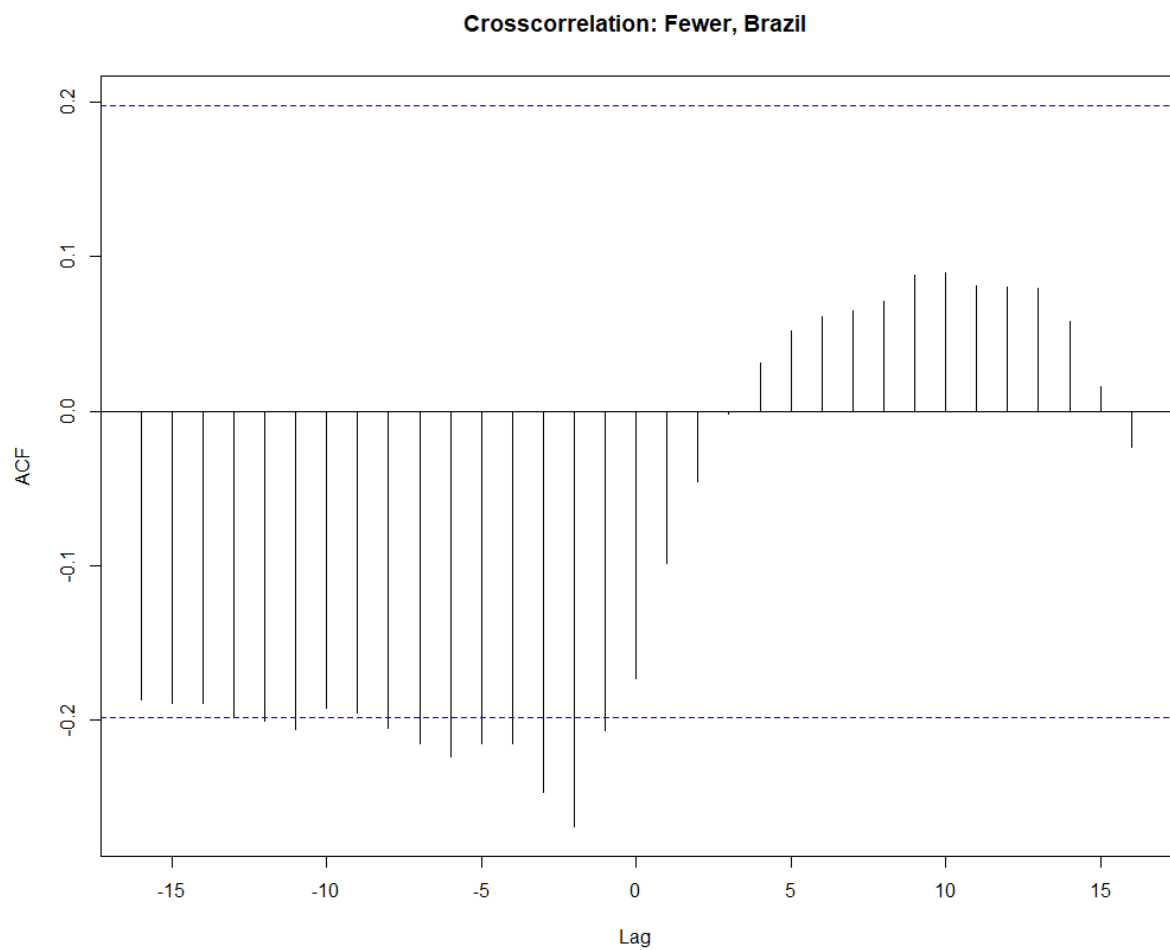


Рисунок 104 График кросскорреляции запросов по теме Лихорадка и заболеваемостью в Бразилии

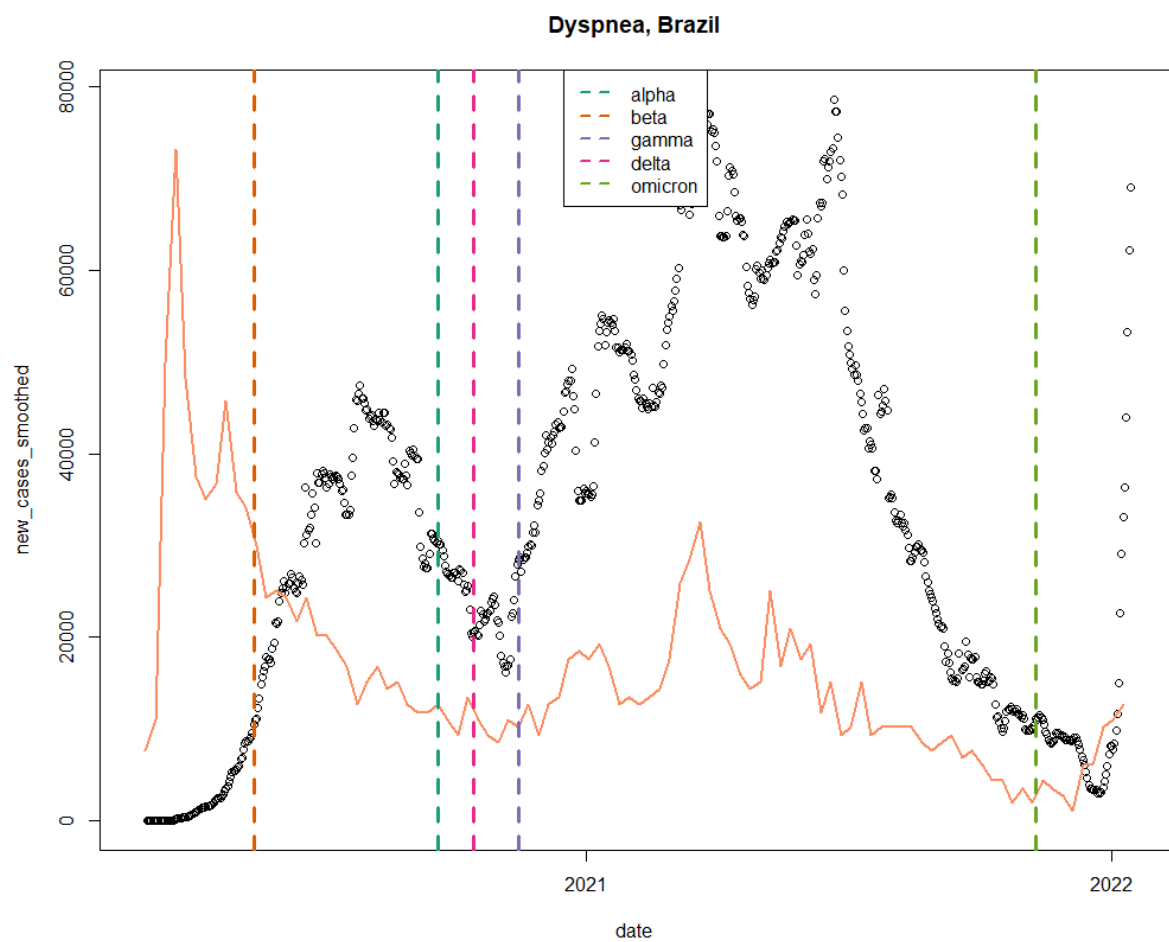


Рисунок 105 Графики популярности запросов по теме Одышка и подтвержденных случаев заболевания в Бразилии

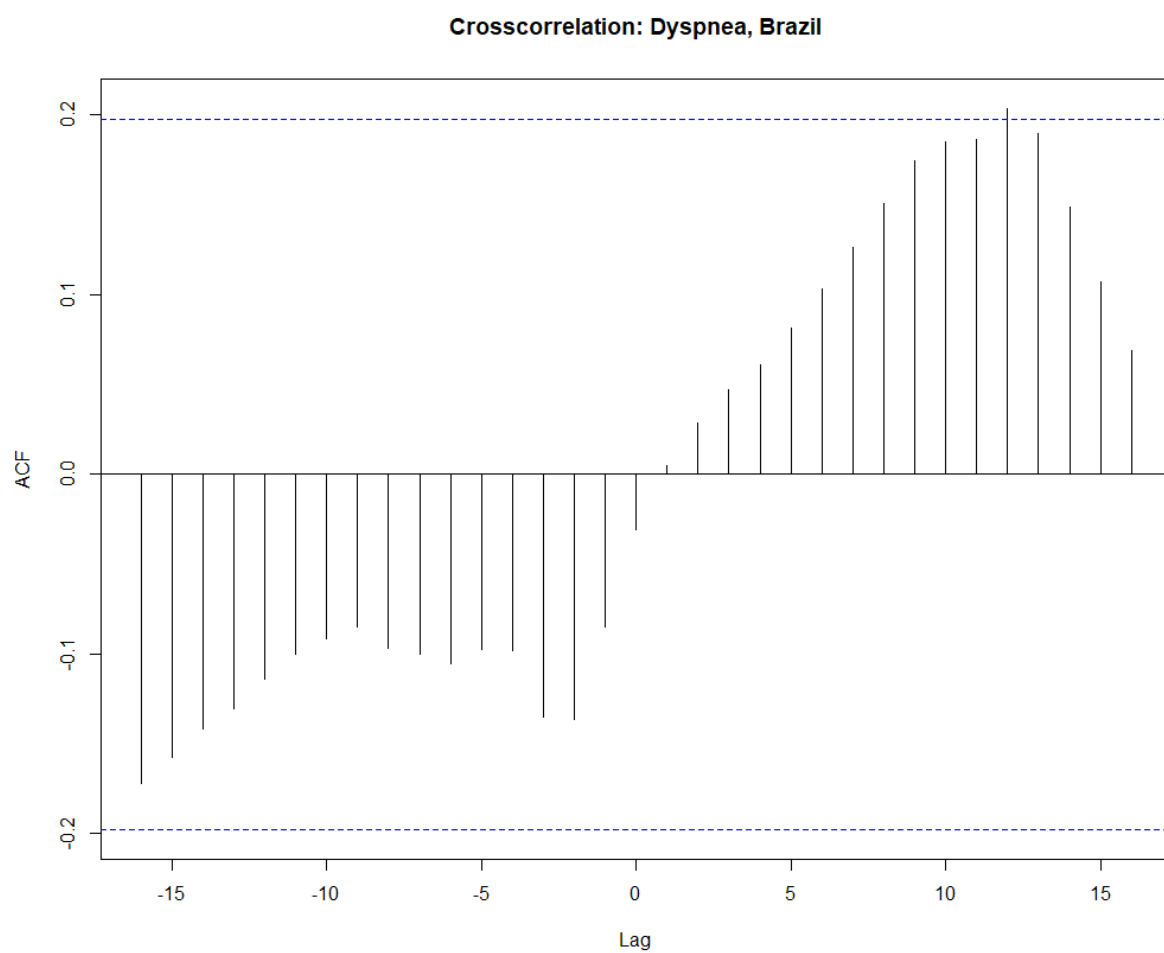


Рисунок 106 График кросскорреляции запросов по теме Одышка и заболеваемостью в Бразилии

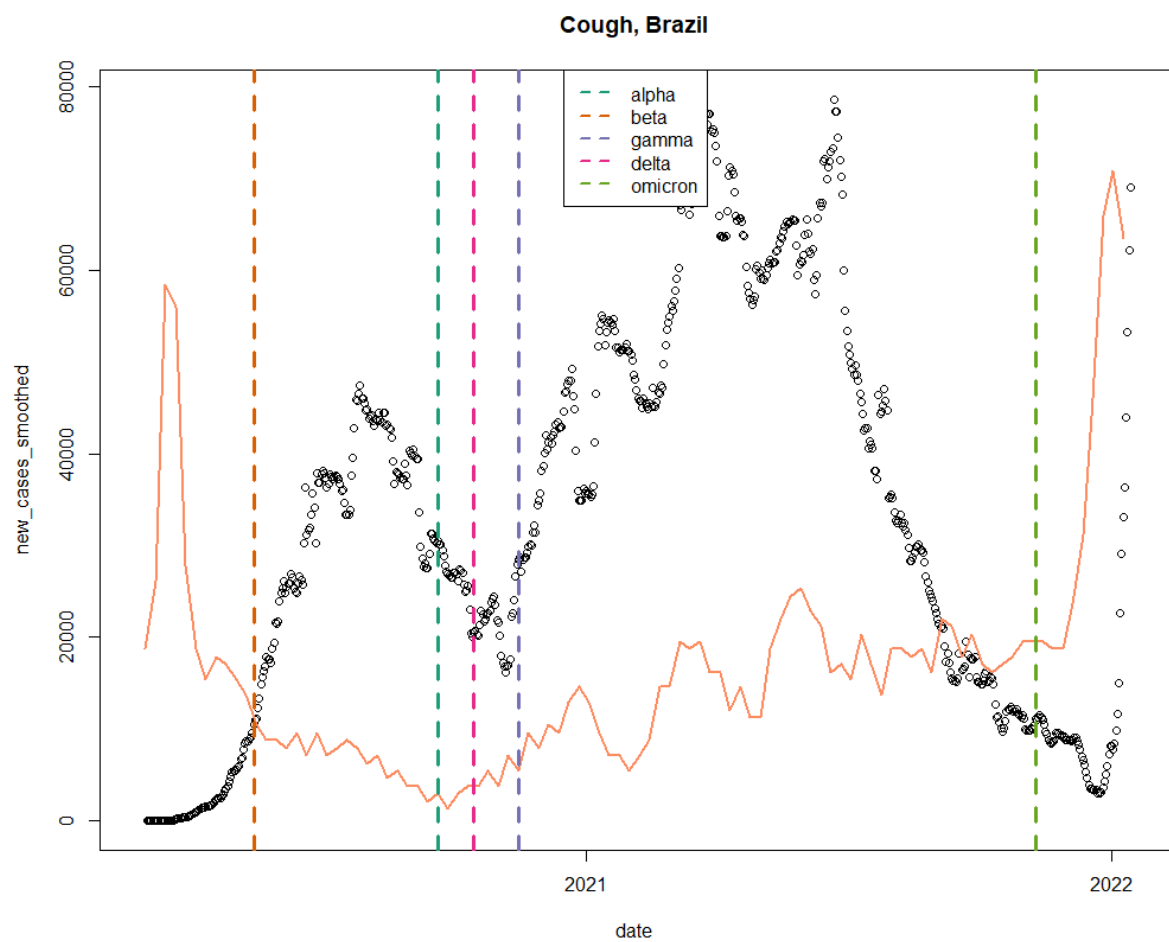


Рисунок 107 Графики популярности запросов по теме Кашель и подтвержденных случаев заболевания в Бразилии

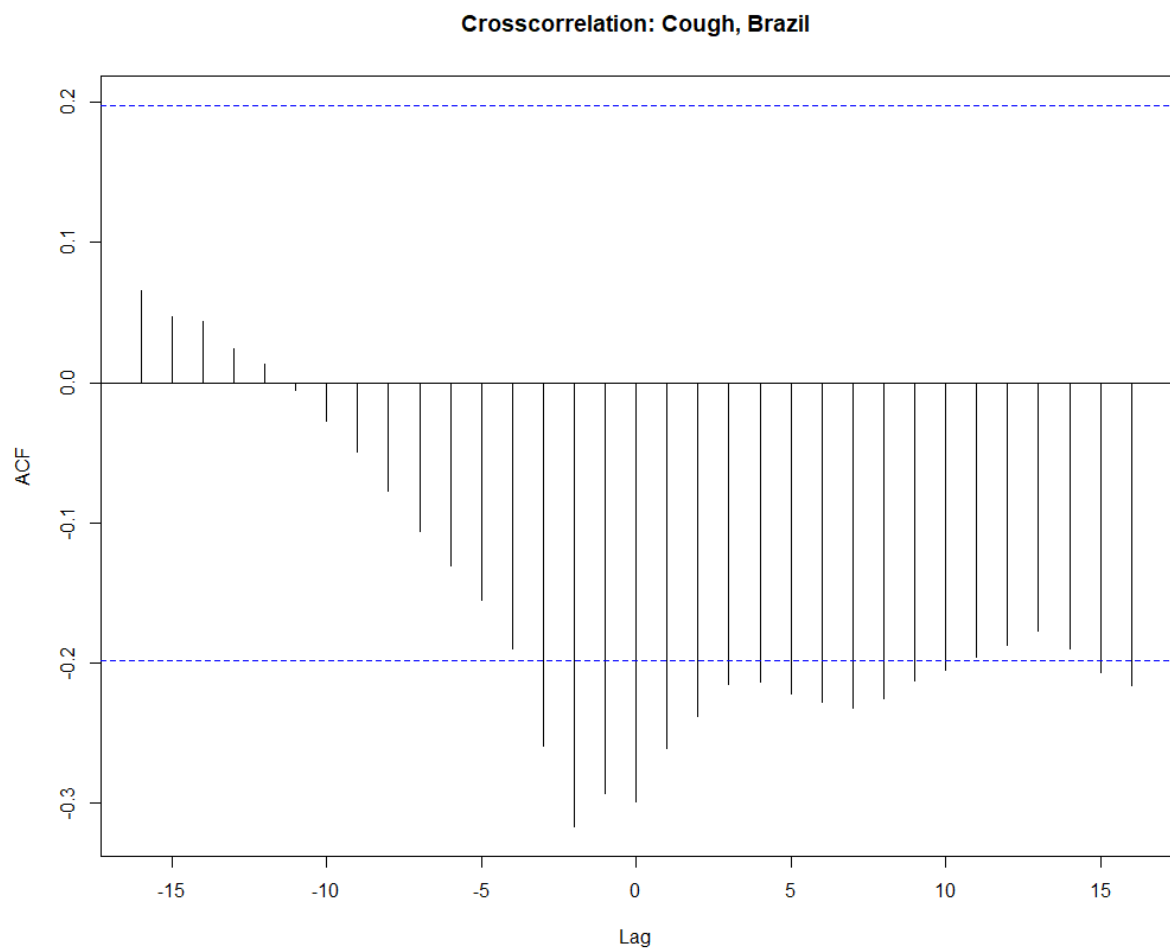


Рисунок 108 График кросскорреляции запросов по теме Кашель и заболеваемостью в Бразилии

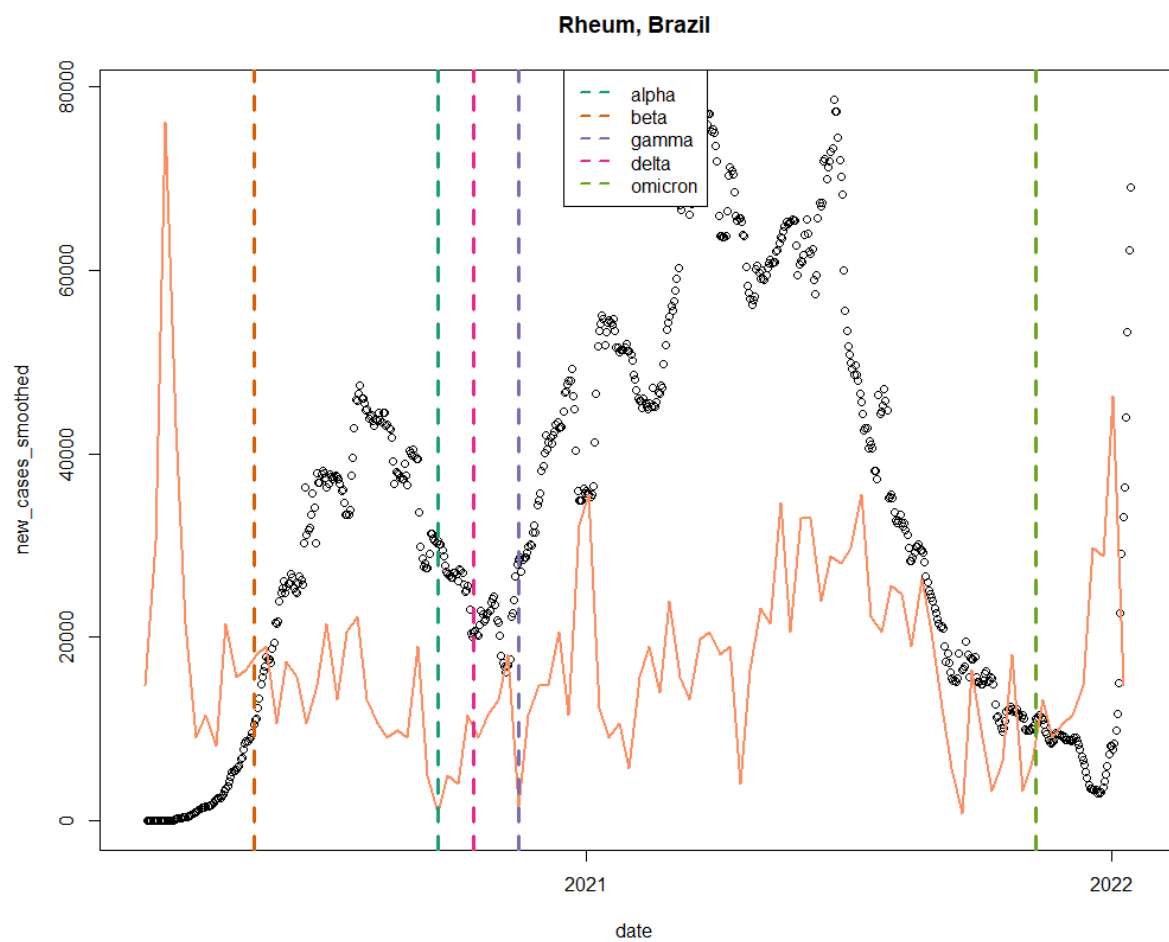


Рисунок 109 Графики популярности запросов по теме Насморк и подтвержденных случаев заболевания в Бразилии

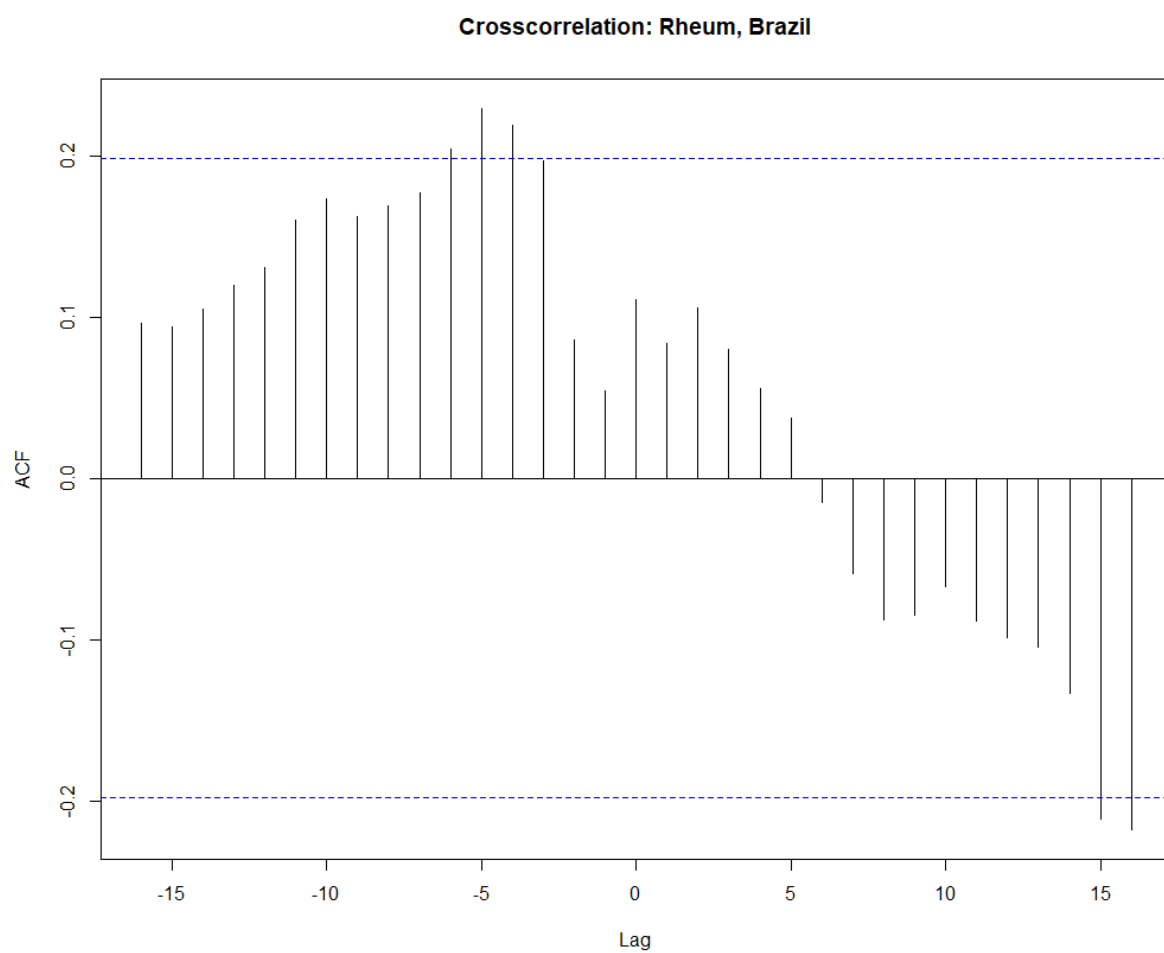


Рисунок 110 График кросскорреляции запросов по теме Насморк и заболеваемостью в Бразилии

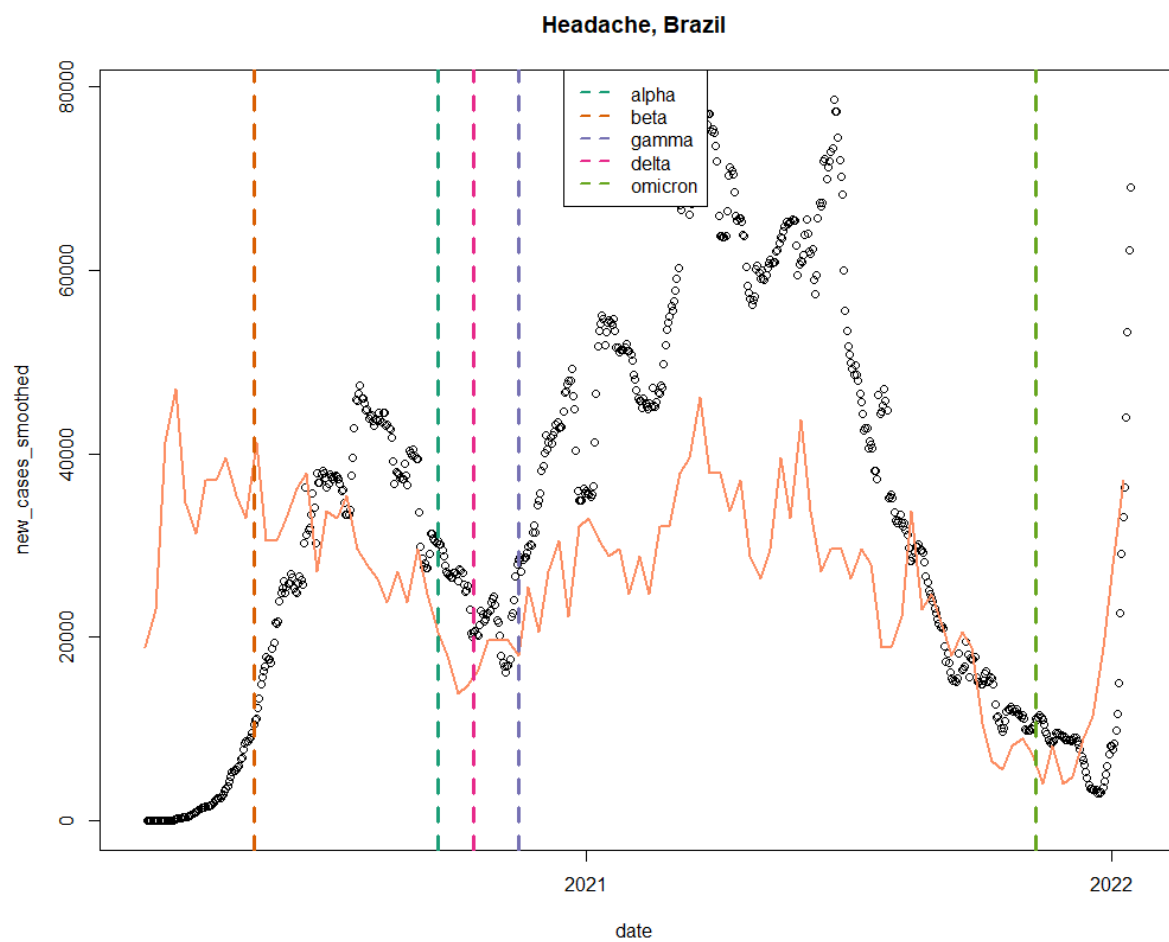


Рисунок 111 Графики популярности запросов по теме Головная боль и подтвержденных случаев заболевания в Бразилии

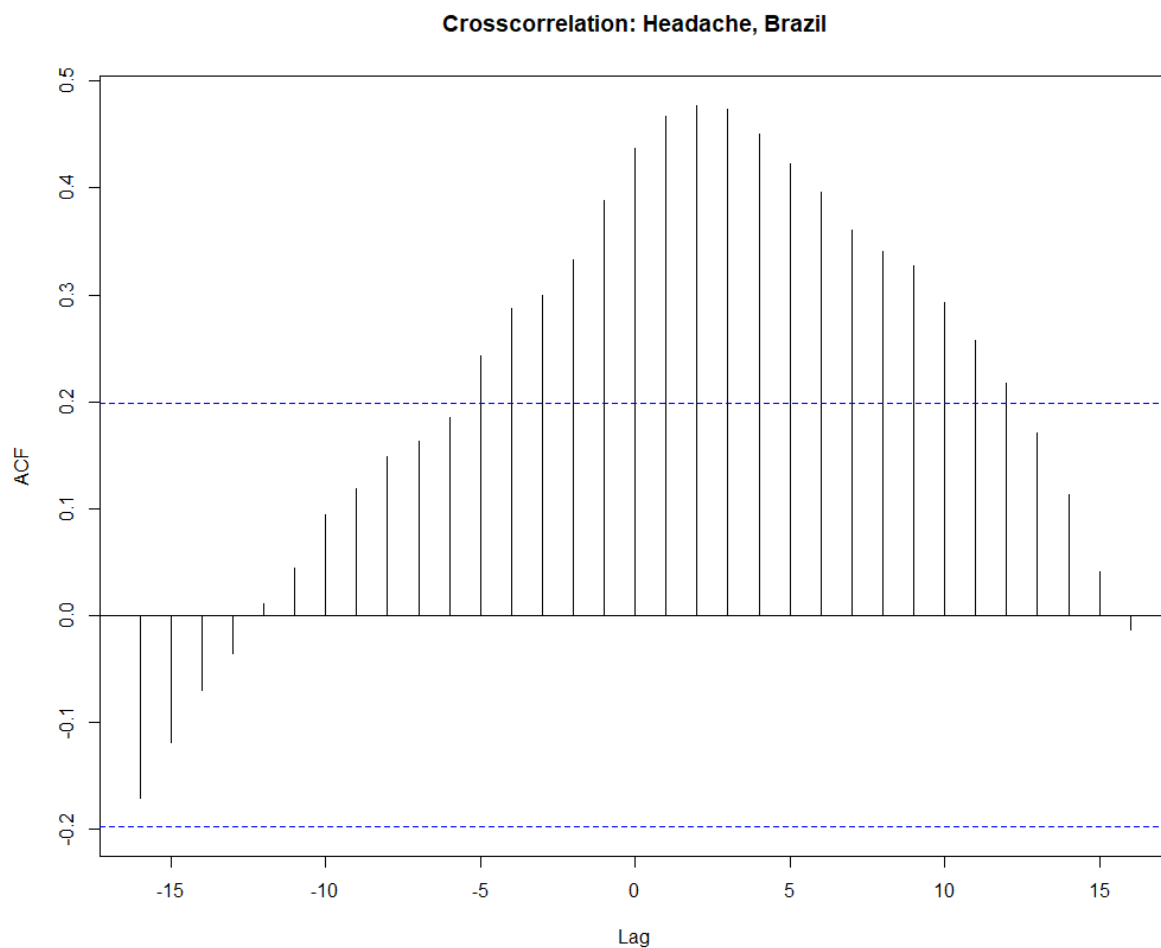


Рисунок 112 График кросскорреляции запросов по теме Головная боль и заболеваемостью в Бразилии

Анализ по стране

Данные по Бразилии отличаются от Великобритании и ЮАР, где рост заболеваемости не сопровождается ростом интереса к большинству запросов. Здесь рост заболеваемости сопровождается ростом интереса ко всем запросам, что может говорить о другом штамме, возможно, о вариации омикрона, т.к. количество заражений растет с такой же прогрессией, как и в ЮАР.

Исследование корреляции ключевых слов по странам:

COVID	PEARSON	SPEARMEN
RUSSIA	0.329257	0.395851
FRANCE	0.469812	0.779743
UNITED STATES	-0.06182	0.119239
INDIA	0.088923	0.207482
UNITED KINGDOM	-0.10118	-0.04721
SOUTH AFRICA	0.592033	0.546078
BRAZIL	-0.15773	0.275985

SMELL	PEARSON	SPEARMEN
RUSSIA	0.396693	0.55849
FRANCE	0.285619	0.638629
UNITED STATES	0.492288	0.597831
INDIA	0.868399	0.770105
UNITED KINGDOM	0.021007	0.299734
SOUTH AFRICA	0.571574	0.541808
BRAZIL	0.298093	0.369626

COUGH	PEARSON	SPEARMEN
RUSSIA	0.340344	0.283081
FRANCE	0.150052	0.170799
UNITED STATES	0.168117	0.072499
INDIA	0.694892	0.143767
UNITED KINGDOM	0.18801	0.478509
SOUTH AFRICA	0.568259	0.400261
BRAZIL	-0.29896	-0.2442

FEVER	PEARSON	SPEARMEN
RUSSIA	0.17487	0.315999
FRANCE	0.084024	0.185965
UNITED STATES	0.183164	0.101951
INDIA	0.858437	0.575025
UNITED KINGDOM	0.024518	0.398593
SOUTH AFRICA	0.311262	0.234764
BRAZIL	-0.17313	0.022171

HEADACHE	PEARSON	SPEARMEN
RUSSIA	0.047515	0.102577
FRANCE	0.071745	0.183587
UNITED STATES	0.324957	0.245963
INDIA	0.775421	0.526135
UNITED KINGDOM	0.002881	0.102502
SOUTH AFRICA	0.379772	0.272044
BRAZIL	0.437231	0.384068

THROAT	PEARSON	SPEARMEN
RUSSIA	-0.11343	-0.13235
FRANCE	0.115282	0.058643
UNITED STATES	0.211912	0.095149
INDIA	0.380654	0.172382
UNITED KINGDOM	0.029718	0.306191
SOUTH AFRICA	0.421824	0.448668
BRAZIL	-0.21112	-0.06871

RHEUM	PEARSON	SPEARMEN
RUSSIA	0.253952	0.256227
FRANCE	0.209374	0.21905
UNITED STATES	0.518913	0.348671
INDIA	0.712237	0.392717
UNITED KINGDOM	-0.07954	0.228139
SOUTH AFRICA	0.155989	0.263895
BRAZIL	0.111077	0.238624

DYSPNEA	PEARSON	SPEARMEN
RUSSIA	0.129586	0.183216
FRANCE	-0.03531	0.012964
UNITED STATES	-0.1734	-0.08512
INDIA	0.634152	0.311559
UNITED KINGDOM	-0.15475	0.016459
SOUTH AFRICA	0.311348	0.269788
BRAZIL	-0.03097	0.293709

Заключение

Взаимосвязь между запросами в интернете, сообщениями в СМИ и фактическими сведениями о заболеваемости зависит от большого числа факторов: было бы полезно наложить даты локдаунов для каждой страны, изучить доступ к интернету у населения и т.д. Тем не менее основные тренды поисковых запросов в интернете в период пандемии могут служить дополнительным компонентом эпидемиологического надзора.

На основе графиков и корреляций мы сделали некоторые эмпирические гипотезы, проверить которые наверняка, скорее всего, не удастся.

Список литературы

1. Weekly epidemiological update on COVID-19 // World Health Organization ([Weekly epidemiological update on COVID-19 - 22 June 2021 \(who.int\)](#))
2. OWID Covid-19 ([Coronavirus Pandemic \(COVID-19\) - Our World in Data](#))
3. Google Trends ([trends.google.ru](#))
4. Симптоматика коронавирусной инфекции // ВОЗ ([Коронавирус \(who.int\)](#))
5. Протокол исследования передачи коронавирусной инфекции 2019 г. // ВОЗ ([WHO-2019-nCoV-HHtransmission-2020.4-rus.pdf](#))
6. Момыналиев К. Т., Акимкин В. Г. Анализ динамики запросов Google Trends в России в период пандемии коронавирусной инфекции как инструмент эпидемиологического надзора. Эпидемиол. инфекц. болезни. Актуал. вопр. 2020; 10
7. Яндекс.Подбор Слов([https://wordstat.yandex.ru/](#))
8. Статистика доли рынка поисковых систем