1. must visit достопримечательности

* создаем новую колонку вместо num of reviews
* отправная точка фильтрации num of reviews: задаем сами с помощью а) summary переменной num of reviews либо б) функция quantile (напр., 10%)

1. кластеризация

* добавить переменную must visit
* исправить код:
* нужно, чтобы были числа, но не так, как мы присваивали, а с помощью функции spread.

Пример, как работает spread:

places$val= 1 – создаем дополнительную колонку

places= places%>%spread (key = “category”, value= “val”, fill=0)

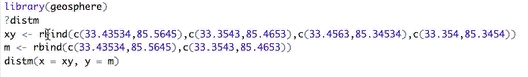
Если функция видит слово “sights”, следовательно в столбце будет записано 1,0.

* все переменные должны быть numeric
* совет: если хотим, чтобы одни переменные учитывались больше, чем другие (напр.duration), то можно домножить – будет больше влиять.

1. расстояние между координатами

* соединить таблицу метро с датасетом достопримечательностей
* код: 2 варианта

1. library (geosphere) – рассчитывает расстояние по прямой



\*расстояние измеряется в метрах

1. open maps – рассчитывает расстояние по дороге
2. стратегии кластеризации – 2 варианта- подумать, какая нам подойдет больше

* анкета с вопросом от общего частного (кластерам присваиваем имена на основе summary of clusters: напр., «долго и пафосно»\ «быстро и дешево» и задаем пользователю вопрос : «Вы хотели бы посетить достопримечательность , которая относится к категории «долго и пафосно»\»дешево и быстро»?
* анкета с вопросом от частного к общему (из каждого кластера выбираем «идеального представителя»-достопримечательность, которая имеет наиболее усредненные параметры по кластеру, плюс она должна быть боле менее «на слуху», напр., Эрмитаж. Так, задаем вопрос пользователю: вам понравилось в Эрмитаже\хотели бы посетить? Если он отвечает да, то можем советовать ему другие достопримечательности из того же кластера, где находится Эрмитаж.

1. Карта

* В лабораторных по digital humanities есть про maps
* АВ обещала что-то еще поискать

1. Критерии оценки рекомендательной системы

* Проверять, насколько результаты рекомендательной системы совпадают с ожиданиями

1. холодный старт

* идея хорошая

1. Неравномерное разбиение на категории

* это нормально

\*Если что, можно дополнить из первоначального датасета другими достопримечательностями с foursquare rating