#Load packages and set directory

library(dplyr)

library(stringr)

library(readxl)

library(reshape2)

setwd("C:/Users/daniel gray/Desktop/CareHome")

#Read in data on cross-setting deaths, average over the last 5 years. Source: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/adhocs/11826fiveyearaverageweeklydeathsbylocalauthorityandplaceofoccurrenceenglandandwalesdeathsregistered2015to2019>

#Paste data from “Five year averages” tab into excel document.

Typical <- read\_excel("weeklyfiveyearaveragesbylaandplaceofoccurrence20152019.xlsx")

#Read in data on cross-setting deaths, this years data. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/datasets/deathregistrationsandoccurrencesbylocalauthorityandhealthboard>

#Copy data from “Occurrences – All data” tab into a new spreadsheet.

ThisYear <- read\_excel("OccurancesAllData.xlsx")

#Read in data on care homes for 2019. Download both the financial file and the SALT file so we can calculate the average spend per patient. <https://digital.nhs.uk/data-and-information/publications/statistical/adult-social-care-activity-and-finance-report/2018-19>

Finance2019 <- read.csv(file="ASC-FRDataFile(withdescriptions).csv", header = TRUE, sep = ",", stringsAsFactors=FALSE)

Clients2019 <- read.csv(file="SALTDataFile(withdescriptions).csv", header = TRUE, sep = ",", stringsAsFactors=FALSE)

#Read in data on care homes for 2019. Download both the financial file and the SALT file so we can calculate numbers. <https://digital.nhs.uk/data-and-information/publications/statistical/adult-social-care-activity-and-finance-report/2017-18>

Finance2018 <- read.csv(file="ASCFRDataFilewithdescriptions2018.csv", header = TRUE, sep = ",", stringsAsFactors=FALSE)

Clients2018 <- read.csv(file="SALTDataFilewithdescriptions2018.csv", header = TRUE, sep = ",", stringsAsFactors=FALSE)

#source of the number of beds in care homes: Care Quality Commission (CQC) <https://www.cqc.org.uk/sites/default/files/HSCA_Active_Locations_01_June_2020.xlsx>

#Save into a file called CareHomeBeds

#Tidy up the CCQ dataset.

BedsRaw <- read\_excel("CareHomeBeds.xlsx")

Beds <- BedsRaw %>% filter(`Location Primary Inspection Category` == "Residential social care") %>% filter(`Care homes beds` >0) %>% select(`Location Local Authority`,`Care homes beds`,`Location Name`, `Location Latest Overall Rating`) %>% rename(DH\_GEOGRAPHY\_NAME = `Location Local Authority`,Beds =`Care homes beds`, Name =`Location Name`, Rating =`Location Latest Overall Rating`)

#Calculate average number of beds in care homes in each local authority

Concentration <- Beds %>% group\_by(DH\_GEOGRAPHY\_NAME) %>% summarise(AverageSize = mean(Beds))

#Calculate average rating of care homes in each local authority

Beds$Rating <- ifelse(Beds$Rating == "Outstanding",4, Beds$Rating)

Beds$Rating <- ifelse(Beds$Rating == "Good",3, Beds$Rating)

Beds$Rating <- ifelse(Beds$Rating == "Requires improvement",2, Beds$Rating)

Beds$Rating <- ifelse(Beds$Rating == "Inadequate",1, Beds$Rating)

Beds$Rating <- as.numeric(Beds$Rating)

Beds$Poor <- ifelse(Beds$Rating <3, 1, 0)

Beds$All <- ifelse(is.na(Beds$Rating),0,1)

Beds$WeightedRating <- Beds$Rating \* Beds$Beds

#Tidy Dataset

Beds <- Beds %>% group\_by(DH\_GEOGRAPHY\_NAME) %>% summarise(Total = sum(Beds,na.rm = TRUE), AverageRating = mean(Rating, na.rm = TRUE), WeightedRating = sum(WeightedRating,na.rm = TRUE) / sum(Beds, na.rm = TRUE))

Beds <- left\_join(Beds,Concentration, by = c("DH\_GEOGRAPHY\_NAME"))

#Rename some local authorities so that they can be matched into other dataset

Beds$DH\_GEOGRAPHY\_NAME <- ifelse(Beds$DH\_GEOGRAPHY\_NAME == "Bournemouth, Christchurch and Poole", "Bournemouth", Beds$DH\_GEOGRAPHY\_NAME)

Beds$DH\_GEOGRAPHY\_NAME <- ifelse(Beds$DH\_GEOGRAPHY\_NAME == "Bristol, City of", "Bristol City of", Beds$DH\_GEOGRAPHY\_NAME)

Beds$DH\_GEOGRAPHY\_NAME <- ifelse(Beds$DH\_GEOGRAPHY\_NAME == "Herefordshire, County of", "Herefordshire County of", Beds$DH\_GEOGRAPHY\_NAME)

Beds$DH\_GEOGRAPHY\_NAME <- ifelse(Beds$DH\_GEOGRAPHY\_NAME == "Kingston upon Hull, City of", "Kingston upon Hull City of", Beds$DH\_GEOGRAPHY\_NAME)

#Reduce down finance datasets. 99 refers to totals for (i) type of support (e.g. mental, physical) and (ii) how that is provided, e.g. under own funds.

Finance2019 <- Finance2019 %>% filter(DimensionGroup == "Expenditure") %>% filter(AgeBand\_Key == "65 and Over") %>% filter(FinanceType\_Key == "Expenditure") %>% filter(CareType\_Key == "Long Term Care") %>% filter(SupportSetting\_Key == "Residential") %>% filter(PrimarySupportReason\_Key == 99) %>% filter (FinanceDescription\_Key == 99) %>% select(GEOGRAPHY\_CODE, DH\_GEOGRAPHY\_NAME,GEOGRAPHY\_NAME, ITEMVALUE) %>% rename(Spending2019 = ITEMVALUE) %>% filter(GEOGRAPHY\_CODE != "")

Finance2018 <- Finance2018 %>% filter(DimensionGroup == "Expenditure") %>% filter(AgeBand\_Key == "65 and Over") %>% filter(FinanceType\_Key == "Expenditure") %>% filter(CareType\_Key == "Long Term Care") %>% filter(SupportSetting\_Key == "Residential") %>% filter(PrimarySupportReason\_Key == 99) %>% filter (FinanceDescription\_Key == 99) %>% select(GEOGRAPHY\_CODE,ITEMVALUE) %>% rename(Spending2018 = ITEMVALUE) %>% filter(GEOGRAPHY\_CODE != "")

#Reduce down clients datasets.

Clients2019 <- Clients2019 %>% filter(TableType\_Key == "Long Term Care") %>% filter(AgeBand\_Key == "65 and over") %>% filter(LongTermSupportSetting\_Key == "Residential") %>% filter(PrimarySupportReason\_Key == 99) %>% filter(LongTermPeriod\_Key == "Year End") %>% select(GEOGRAPHY\_CODE, ITEMVALUE) %>% rename(Clients2019 = ITEMVALUE) %>% filter(GEOGRAPHY\_CODE != "")

Clients2018 <- Clients2018 %>% filter(TableType\_Key == "Long Term Care") %>% filter(AgeBand\_Key == "65 and over") %>% filter(LongTermSupportSetting\_Key == "Residential") %>% filter(PrimarySupportReason\_Key == 99) %>% filter(LongTermPeriod\_Key == "Year End") %>% select(GEOGRAPHY\_CODE, ITEMVALUE) %>% rename(Clients2018 = ITEMVALUE) %>% filter(GEOGRAPHY\_CODE != "")

#Create consolidated dataset

Summary <- left\_join(Finance2019,Finance2018, by = c("GEOGRAPHY\_CODE"))

Summary <- left\_join(Summary,Clients2019, by = c("GEOGRAPHY\_CODE"))

Summary <- left\_join(Summary,Clients2018, by = c("GEOGRAPHY\_CODE"))

Summary <- left\_join(Summary,Beds, by = c("DH\_GEOGRAPHY\_NAME"))

rm(Clients2018,Clients2019,Finance2019,Finance2018,Beds)

#Work out when the death rate exceeded 20 in each local authority (as a proxy for when outbreak first took grip of an area)

CommunityDeath <- ThisYear %>% filter(`Cause of death` == "COVID 19") %>% group\_by(`Area code`,`Week number`) %>% mutate(Total = sum(`Number of deaths`)) %>% distinct(`Area code`,`Week number`, Total, `Area name`) %>% group\_by(`Area code`) %>% arrange(`Week number`) %>% mutate(cumcases = cumsum(`Total`)) %>% filter(cumcases > 20) %>% mutate(Community = min(`Week number`)) %>% distinct(`Area code`, Community) %>% rename(GEOGRAPHY\_CODE = `Area code`)

#Work out care home deaths and non-care home deaths.

CareHomes <- ThisYear %>% filter(`Cause of death` == "All causes") %>% filter(`Place of death` == "Care home") %>% group\_by(`Area code`) %>% mutate(CareHomeDeaths2020 = sum(`Number of deaths`)) %>% filter(`Week number` == max(`Week number`)) %>% select(`Area code`, `CareHomeDeaths2020`,`Area name`) %>% rename(GEOGRAPHY\_CODE = `Area code`, LocalAuthority = `Area name`)

Other <- ThisYear %>% filter(`Cause of death` == "All causes") %>% filter(`Place of death` != "Care home") %>% group\_by(`Area code`) %>% mutate(OtherDeaths2020 = sum(`Number of deaths`)) %>% filter(`Week number` == max(`Week number`)) %>% select(`Area code`, `OtherDeaths2020`) %>% distinct(`Area code`, `OtherDeaths2020`) %>% rename(GEOGRAPHY\_CODE = `Area code`)

#Create England only dataset for aggregate figures.

England <- left\_join(CareHomes,Other, by = c("GEOGRAPHY\_CODE"))

#Consolidate into main dataset

CareHomes <- CareHomes %>% select(GEOGRAPHY\_CODE, CareHomeDeaths2020)

Summary <- left\_join(Summary,CareHomes, by = c("GEOGRAPHY\_CODE"))

Summary <- left\_join(Summary,Other, by = c("GEOGRAPHY\_CODE"))

Summary <- left\_join(Summary,CommunityDeath, by = c("GEOGRAPHY\_CODE"))

#Work out care home deaths and non-care home deaths typical year

CareHomes <- Typical %>% filter(`Place of occurrence` == "Care home") %>% group\_by(`Local Authority Code`) %>% filter(`Week Number` < 22) %>% mutate(CareHomesTypical = sum(`Five year average number of deaths`)) %>% select(`Local Authority Code`, `CareHomesTypical`) %>% distinct(`Local Authority Code`, `CareHomesTypical`) %>% rename(GEOGRAPHY\_CODE = `Local Authority Code`)

Other <- Typical %>% filter(`Place of occurrence` != "Care home") %>% group\_by(`Local Authority Code`) %>% filter(`Week Number` < 22) %>% mutate(OtherDeathsTypical = sum(`Five year average number of deaths`)) %>% select(`Local Authority Code`, `OtherDeathsTypical`) %>% distinct(`Local Authority Code`, `OtherDeathsTypical`) %>% rename(GEOGRAPHY\_CODE = `Local Authority Code`)

#Consolidate into main dataset.

Summary <- left\_join(Summary,CareHomes, by = c("GEOGRAPHY\_CODE"))

Summary <- left\_join(Summary,Other, by = c("GEOGRAPHY\_CODE"))

#Work out total deaths in England.

England <- left\_join(England,Other, by = c("GEOGRAPHY\_CODE"))

England <- left\_join(England,CareHomes, by = c("GEOGRAPHY\_CODE"))

England$Wales <- ifelse(str\_sub(England$GEOGRAPHY\_CODE,1,1) == "W", "Wales","England")

England <- England %>% filter(Wales == "England")

England$ExcessCareHome <- England$CareHomeDeaths2020 - England$CareHomesTypical

England$ExcessCareHome <- ifelse(England$ExcessCareHome <0, 0, England$ExcessCareHome)

Fact1 <- sum(England$ExcessCareHome, na.rm = TRUE)

Fact2 <- sum(England$CareHomesTypical, na.rm = TRUE)

BedsNames <- Summary %>% select(GEOGRAPHY\_CODE, Total)

Fact3 <- sum(BedsNames$Total, na.rm = TRUE)

Fact4 <- Fact1 / Fact3

England <- left\_join(England,BedsNames, by = c("GEOGRAPHY\_CODE"))

#Only includes data for metropolitan districts, London boroughs and unitary authorities. So doesn’t include district councils. Therefore remove when calculating average impact.

England <- na.omit(England)

#Create facts for article.

Fact5 <- sum(England$ExcessCareHome) / sum(England$Total)

#Tidy up main dataset.

Summary <- na.omit(Summary)

Summary$Spending2019 <- as.numeric(Summary$Spending2019)

Summary$Clients2019 <- as.numeric(Summary$Clients2019)

Summary$Spending2018 <- as.numeric(Summary$Spending2018)

Summary$Clients2018 <- as.numeric(Summary$Clients2018)

#Create new fields: Spend per patient.

Summary$Spend <- Summary$Spending2019 / Summary$Clients2019

Summary$Spend2018 <- Summary$Spending2018 / Summary$Clients2018

#Create new fields: Excess Death Rate

Summary$ExcessShare <- (Summary$CareHomeDeaths2020 - Summary$CareHomesTypical) / Summary$Total

Summary$Other <- (Summary$OtherDeaths2020 - Summary$OtherDeathsTypical) / Summary$OtherDeathsTypical

Summary$CareHomeNormal <- Summary$CareHomesTypical / Summary$Total

#Create new fields: Spend compared to regional average

Summary <- Summary %>% group\_by(GEOGRAPHY\_NAME) %>% mutate(RegionalSpending = sum(Spending2019), RegionalClients = sum(Clients2019))

Summary$RegionalSpend <- Summary$RegionalSpending / Summary$RegionalClients

Summary$Diff <- (Summary$Spend - Summary$RegionalSpend)

#Tidy: remove Slough as numbers look very weird.

Summary <- Summary %>% filter (DH\_GEOGRAPHY\_NAME != "Slough")

Summary <- na.omit(Summary)

#Create spreadsheet for sharing

Summary <- Summary %>% select(-c("Spending2018","Spend2018", "AverageRating", "Clients2018","Spending2019","Clients2019","RegionalSpending","RegionalClients"))

Summary <- Summary %>% rename(Region = GEOGRAPHY\_NAME, Total.Beds = Total, AverageRating = WeightedRating, AverageBedsPerCareHome = AverageSize, DifferanceFromRegional = Diff, ExcessCareHome = ExcessShare, ExcessOther = Other)

#Create spreadsheet for sharing

Chart <- Summary %>% arrange(desc(ExcessCareHome)) %>% head (n=20)

#Create final fact

Brent <- Summary %>% filter(DH\_GEOGRAPHY\_NAME == "Brent")

Brent <- Brent$ExcessCareHome

Hammersmith <- Summary %>% filter(DH\_GEOGRAPHY\_NAME == "Hammersmith and Fulham")

Hammersmith <- Hammersmith$ExcessCareHome

Fact6 <- Brent/Hammersmith

Fact7 <- Summary %>% ungroup() %>% summarise(CountryAverageRating = mean(AverageRating), CountryAverageSize = mean(AverageBedsPerCareHome), CountryExcessOther = mean(ExcessOther))

#Check regression gives similar intuition to main narrative.

Regression <- Summary %>% filter(Other > -0.01)

Model <- lm(ExcessShare ~ Other + WeightedRating + AverageSize + Community + Diff, data = Regression)

summary(Model)