

# Title Placeholder

## Load in Specific Packages

```
library(readr)
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(here)
```

```
## here() starts at C:/Users/romin/ToyRepo
```

```
library(lemon)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##   group_rows
```

```
library(ggplot2)
library(reshape())
```

```
##
## Attaching package: 'reshape'

## The following object is masked from 'package:dplyr':
##
##   rename

## The following objects are masked from 'package:tidyr':
##
##   expand, smiths
```

```
library(hexbin)
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following object is masked from 'package:reshape':
```

```
##
##      melt
## The following objects are masked from 'package:dplyr':
##
##      between, first, last
library(GGally)

## Registered S3 method overwritten by 'GGally':
##      method from
##      +.gg      ggplot2
library(formattable)
library(viridis)

## Loading required package: viridisLite
library(TTR)
library(zoo)

##
## Attaching package: 'zoo'
## The following objects are masked from 'package:data.table':
##
##      yearmon, yearqtr
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
library(ggrepel)
library(grid)
```

## Load in the data

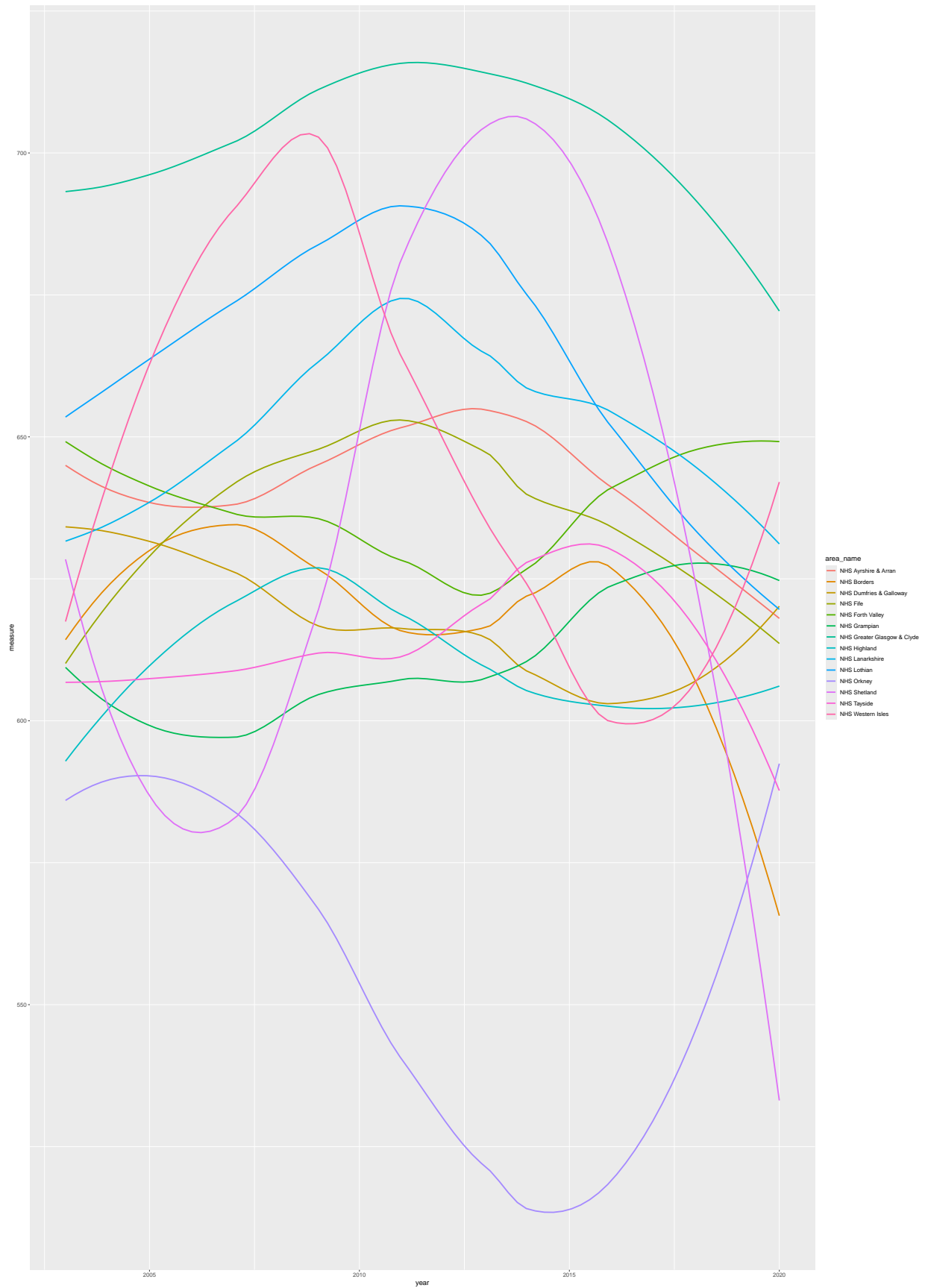
```
cancerReg <- read.csv("C:\\Users\\romin\\ToyRepo\\Models\\cancerReg.csv")
```

## Remove Unnecessary Data for Analysis

```
cancerReg <- cancerReg %>% select(-period, -area_type, -type_definition, -indicator, -upper_confidence_int)
```

## Display All Data Points

```
## `geom_smooth()` using formula = 'y ~ x'
```



## Find Average of All Measures by Year

```
avgYearly <- cancerReg %>%  
  group_by(year) %>%  
  mutate(AvgYear = mean(measure, na.rm = TRUE)) %>%  
  select(-area_name, -measure, -area_code)
```

## Calculate Moving Average for Each Health Board

```
movingAvg <- cancerReg %>%  
  group_by(area_name) %>%  
  arrange(year) %>%  
  mutate(MA = cumsum(measure) / row_number())
```

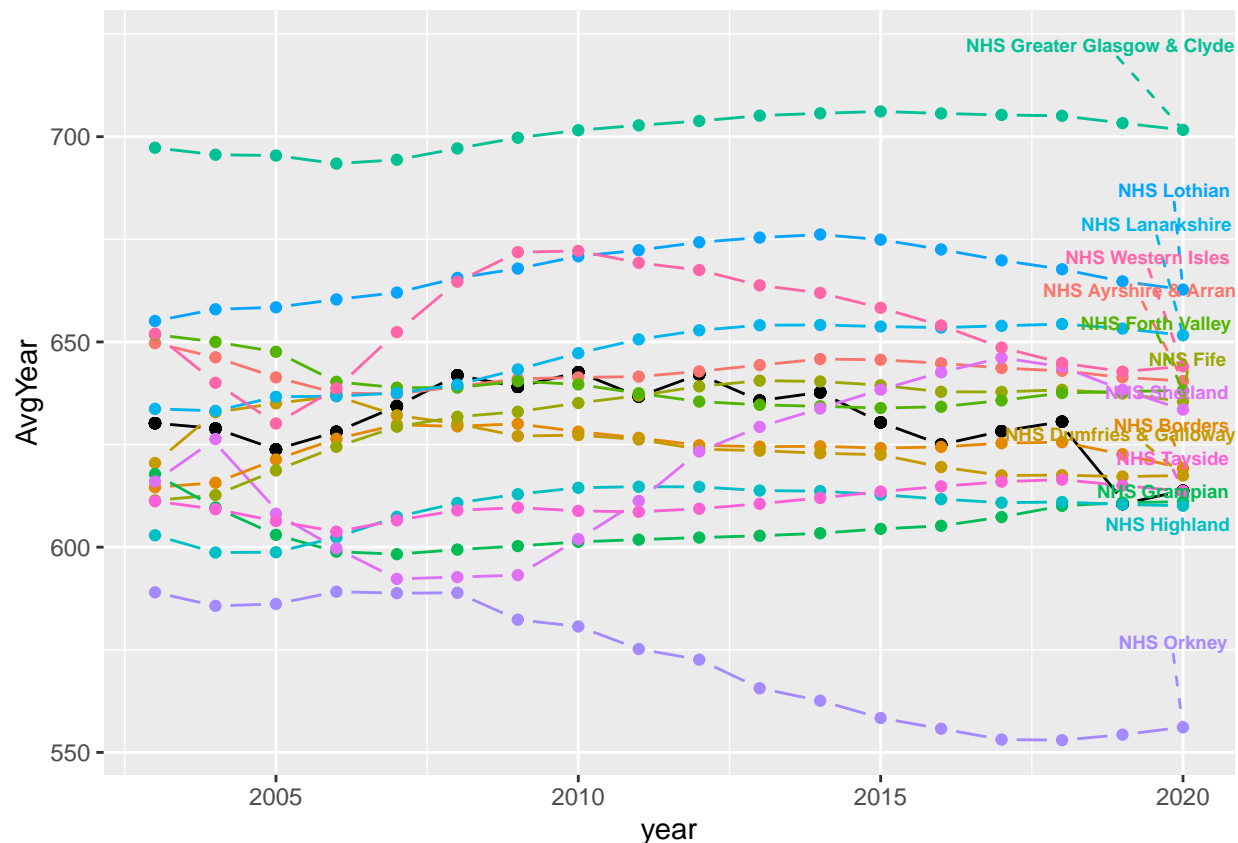
## Find Last Data Points for Data

```
finalValues <- movingAvg %>%  
  group_by(area_name) %>%  
  summarise(  
    lastMA = dplyr::last(MA),  
    lastYear=dplyr::last(year)  
  )
```

## Display Summary of All Data

```
ggplot(data=cancerReg, aes(x = year)) +  
  geom_pointline(data= avgYearly, aes(y = AvgYear)) +  
  geom_pointline(data=movingAvg, aes(y=MA, col=area_name)) +  
  geom_text_repel(data = finalValues, aes(  
    x = lastYear,  
    y = lastMA,  
    label = area_name,  
    color=area_name),  
    size = 2.5,  
    fontface = "bold",  
    nudge_y = 20.6,  
    direction = "y",  
    hjust= -0.7,  
    segment.linetype=2,  
    segment.size = 0.5,  
    segment.curvature=0  
  ) +  
  theme(legend.position = "none")
```

```
## Warning in geom_pointline(data = avgYearly, aes(y = AvgYear)): `geom_pointpath`  
## and `geom_pointline` have been soft-deprecated. A replacement can be found in  
## ggh4x::geom_pointpath.  
## Warning in geom_pointline(data = movingAvg, aes(y = MA, col = area_name)):  
## `geom_pointpath` and `geom_pointline` have been soft-deprecated. A replacement  
## can be found in ggh4x::geom_pointpath.
```



### Calculate Differences Function

```
sigPercent <- data.frame(
  area_name = character(),
  year = integer(),
  percentNum= numeric(),
  stringsAsFactors = FALSE
)

boardAvg <- function(currBoard, currVal,currYear){

  currAvgYear <- filter(movingAvg, area_name == currBoard & year == currYear) %>% select(MA)
  numCurrAvgYear <- gsub("[^0-9.]", "", currAvgYear$MA)
  numCurrAvgYear <- as.numeric(numCurrAvgYear)
  diffVal <- currVal - numCurrAvgYear
  percentVal <- ((diffVal / numCurrAvgYear) * 100)

  if (percentVal >=3 || percentVal <=-3){
    sigPercent <- sigPercent %>% add_row(area_name = currBoard, year = currYear, percentNum = round(
  )
  }
  return(sigPercent)

## Turn these differences into a percentage of how different it is, and if its greatly different return
# While the moving average does not provide a direct estimation of the predicted values it still serve
}
```

### Calculate Differences

```
healthBoards <- unique(cancerReg$area_name)
totalYears <- unique(cancerReg$year)
for (currBoard in healthBoards) {
  for (currYear in totalYears) {
    currVal <- subset(cancerReg, year == currYear & area_name == currBoard)
    currVal <- select(currVal, -area_code, -area_name, -year)
    currVal <- as.numeric(currVal)
    sigPercent<- boardAvg(currBoard, currVal, currYear)
  }
}
```

[illegible]

[illegible]





[illegible]

[illegible]

```
##          area_name year percentNum
## 1 NHS Borders 2020          -9.54
```

## 2	NHS Lothian 2020	-5.11
## 3	NHS Orkney 2020	5.48
## 4	NHS Shetland 2020	-12.88
## 5	NHS Western Isles 2020	3.38
## 6	NHS Fife 2020	-5.27
## 7	NHS Tayside 2020	-3.77
## 8	NHS Greater Glasgow & Clyde 2020	-4.01
## 9	NHS Lanarkshire 2020	-4.22
## 10	NHS Ayrshire & Arran 2019	-3.96
## 11	NHS Borders 2019	-7.56
## 12	NHS Lothian 2019	-7.11
## 13	NHS Orkney 2019	3.89
## 14	NHS Shetland 2019	-14.09
## 15	NHS Western Isles 2019	-5.25
## 16	NHS Tayside 2019	-3.74
## 17	NHS Greater Glasgow & Clyde 2019	-4.05
## 18	NHS Forth Valley 2018	4.26
## 19	NHS Grampian 2018	6.72
## 20	NHS Lothian 2018	-4.85
## 21	NHS Shetland 2018	-5.15
## 22	NHS Western Isles 2018	-8.71
## 23	NHS Dumfries & Galloway 2017	-4.66
## 24	NHS Forth Valley 2017	3.40
## 25	NHS Grampian 2017	4.95
## 26	NHS Lothian 2017	-5.55
## 27	NHS Orkney 2017	-6.66
## 28	NHS Shetland 2017	7.74
## 29	NHS Western Isles 2017	-11.61
## 30	NHS Dumfries & Galloway 2016	-6.24
## 31	NHS Lothian 2016	-4.63
## 32	NHS Orkney 2016	-6.13
## 33	NHS Shetland 2016	8.53
## 34	NHS Western Isles 2016	-8.55
## 35	NHS Fife 2016	-3.25
## 36	NHS Orkney 2015	-9.02
## 37	NHS Shetland 2015	8.62
## 38	NHS Western Isles 2015	-6.64
## 39	NHS Tayside 2015	3.02
## 40	NHS Orkney 2014	-5.95
## 41	NHS Shetland 2014	7.78
## 42	NHS Western Isles 2014	-3.04
## 43	NHS Orkney 2013	-12.29
## 44	NHS Shetland 2013	9.53
## 45	NHS Western Isles 2013	-5.60
## 46	NHS Dumfries & Galloway 2012	-3.39
## 47	NHS Orkney 2012	-4.10
## 48	NHS Shetland 2012	17.39
## 49	NHS Fife 2012	3.17
## 50	NHS Lanarkshire 2012	3.04
## 51	NHS Orkney 2011	-7.63
## 52	NHS Shetland 2011	12.16
## 53	NHS Western Isles 2011	-3.48
## 54	NHS Lanarkshire 2011	4.13
## 55	NHS Lothian 2010	3.12

```
## 56          NHS Shetland 2010      10.21
## 57          NHS Lanarkshire 2010     4.29
## 58              NHS Orkney 2009    -6.75
## 59          NHS Western Isles 2009   6.42
## 60          NHS Lanarkshire 2009   3.47
## 61          NHS Western Isles 2008   9.23
## 62          NHS Highland 2007      3.29
## 63          NHS Shetland 2007     -5.01
## 64          NHS Western Isles 2007   8.40
## 65              NHS Fife 2007      3.11
## 66          NHS Forth Valley 2006  -3.43
## 67          NHS Shetland 2006     -4.24
## 68          NHS Western Isles 2006   4.02
## 69          NHS Shetland 2005     -5.98
## 70          NHS Western Isles 2005  -3.15
```

```
colourCells <- function(values, average){
  diffVal <- values - average
  if (diffVal>=0 & diffVal<=20){
    return(paste0("\\cellcolor{green!," ,round(diffVal/100), "}"))
  }
}
```

## Summary Table of Data Within Graph

```
inputFile <- "reportReg.pdf"
# healthBoardData <- cancerReg %>%
#   group_by(year) %>%
#   mutate(Percent = measure / sum(measure) * 100)
# healthBoardData <- cancerReg %>% pivot_wider(names_from = year, values_from = measure)
# healthBoardData <- healthBoardData %>%
#   mutate(
#     Average = rowMeans(select(., `2003`:`2020`), na.rm = TRUE)
#   )

sigPercentWide <- sigPercent %>% pivot_wider(
  names_from = area_name,
  values_from = percentNum
)
print(sigPercentWide)
```

```
## # A tibble: 16 x 5
##   year `NHS Borders` `NHS Lothian` `NHS Orkney` `NHS Shetland`
##   <int>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 2020     -9.54      -5.11       5.48     -12.9
## 2 2019     -7.56      -7.11       3.89     -14.1
## 3 2018      NA      -4.85       NA       -5.15
## 4 2017      NA     -5.55      -6.66       7.74
## 5 2016      NA     -4.63      -6.13       8.53
## 6 2015      NA      NA      -9.02       8.62
## 7 2014      NA      NA     -5.95       7.78
## 8 2013      NA      NA    -12.3       9.53
## 9 2012      NA      NA     -4.1      17.4
## 10 2011      NA      NA    -7.63      12.2
```

```
## 11 2010      NA      3.12      NA      10.2
## 12 2009      NA      NA      -6.75      NA
## 13 2008      NA      NA      NA      NA
## 14 2007      NA      NA      NA      -5.01
## 15 2006      NA      NA      NA      -4.24
## 16 2005      NA      NA      NA      -5.98
## # i 10 more variables: `NHS Western Isles` <dbl>, `NHS Fife` <dbl>,
## #   `NHS Tayside` <dbl>, `NHS Greater Glasgow & Clyde` <dbl>,
## #   `NHS Lanarkshire` <dbl>, `NHS Ayrshire & Arran` <dbl>,
## #   `NHS Forth Valley` <dbl>, `NHS Grampian` <dbl>,
## #   `NHS Dumfries & Galloway` <dbl>, `NHS Highland` <dbl>
```

```
kable(sigPercentWide, format = "latex", booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped", "hold_poistion")) %>%
  row_spec(0, bold = TRUE) %>%
  kableExtra::landscape()
```

year	NHS Borders	NHS Lothian	NHS Orkney	NHS Shetland	NHS Western Isles	NHS Fife	NHS Tayside	NHS Greater Glasgow & Clyde
2020	-9.54	-5.11	5.48	-12.88	3.38	-5.27	-3.77	
2019	-7.56	-7.11	3.89	-14.09	-5.25	NA	-3.74	
2018	NA	-4.85	NA	-5.15	-8.71	NA	NA	
2017	NA	-5.55	-6.66	7.74	-11.61	NA	NA	
2016	NA	-4.63	-6.13	8.53	-8.55	-3.25	NA	
2015	NA	NA	-9.02	8.62	-6.64	NA	3.02	
2014	NA	NA	-5.95	7.78	-3.04	NA	NA	
2013	NA	NA	-12.29	9.53	-5.60	NA	NA	
2012	NA	NA	-4.10	17.39	NA	3.17	NA	
2011	NA	NA	-7.63	12.16	-3.48	NA	NA	
2010	NA	3.12	NA	10.21	NA	NA	NA	
2009	NA	NA	-6.75	NA	6.42	NA	NA	
2008	NA	NA	NA	NA	9.23	NA	NA	
2007	NA	NA	NA	-5.01	8.40	3.11	NA	
2006	NA	NA	NA	-4.24	4.02	NA	NA	
2005	NA	NA	NA	-5.98	-3.15	NA	NA	

```
qpdf::pdf_rotate_pages(inputFile, pages=4 , angle=90)
```

```
## [1] "C:\\Users\\romin\\ToyRepo\\Models\\reportReg_output.pdf"
```

#Note for next time: what I want to do at this point is to show the changing colours as a difference change if its only within a small amount of chaning values then ignore the calues and do not #colour the cell, otherwise red fir a rise and green for a fall