

# Behaviour Change: Lockdown Activities

Lara Ollerenshaw

22/04/2021

## Data Origins

The data I have chosen to use for my project is from the *Office of National Statistics* and was collected through an online time-use study. The full data set shows how much time people spent doing various activities during 2014/15 and March/April 2020 and has been broken down to account for income, age, gender, and with/without children. For my project, I focused specifically on the data categorized by income.

## Research Question

Compared to 2014/2015, how did the behavior of the UK population change as a result of lockdown?

Did the change in behavior as a result of lockdown differ between different income brackets?

## Data preparation

Below are the first few rows of the raw data, before removing unwanted columns and relabeling columns to make them meaningful:

```
##                                     X
## 1
## 2
## 3 Travelling and transport (e.g. walking or driving)
## 4                               Working not from home
## 5                               Working from home
## 6                               Study
##                               X2014.2015
## 1 Average daily time (minutes) split by income bands
## 2                               Up to £1,700 per month (a1)
## 3                               65.2
## 4                               70.3
## 5                               6.4
## 6                               16.6
##                               X.1                               X.2
## 1
## 2 £1,700 to £3,300 per month (b1) Over £3300 per month (c1)
## 3                               84.6                               102.1
## 4                               173.7                               202.0
## 5                               11.4                               24.0
## 6                               13.9                               13.6
##                               X.3
## 1 Significant differences between minutes spent by income bands
## 2                               Up to £1,700 per month
## 3                               b1 c1
## 4                               b1 c1
```

```

## 5
## 6
##
##          X.4          X.5
## 1
## 2 £1,700 to £3,300 per month Over £3300 per month
## 3          a1 c1          a1 b1
## 4          a1 c1          a1 b1
## 5          a1 c1          a1 b1
## 6
##
##          March.April.2020
## 1 Average daily time (minutes) split by income bands
## 2          Up to £1,700 per month (a2)
## 3
## 4          21.6
## 5          65.7
## 6          32.4
## 7          11.7
##
##          X.6          X.7
## 1
## 2 £1,700 to £3,300 per month (b2) Over £3300 per month (c2)
## 3          16.9          17.1
## 4          102.1          138.1
## 5          51.1          89.3
## 6          5.7          5.9
##
##          X.8
## 1 Significant differences between minutes spent by income bands
## 2          Up to £1,700 per month
## 3
## 4          c2
## 5          c2
## 6
##
##          X.9          X.10
## 1
## 2 £1,700 to £3,300 per month Over £3300 per month
## 3
## 4          a2
## 5          c2          a2 b2
## 6
##
##          X2014.2015.vs.March.April.2020
## 1 Significant differences between 2014/2015 and March/April 2020
## 2          Up to £1,700 per month
## 3          Significant change
## 4          No Change
## 5          Significant change
## 6          No Change
##
##          X.11          X.12
## 1
## 2 £1,700 to £3,300 per month Over £3300 per month
## 3          Significant change Significant change
## 4          Significant change Significant change
## 5          Significant change Significant change
## 6          Significant change No Change

```

The following shows the first steps that were taken to make the data easier to manage.

```

# Load the data on income
df1_orig <- read.csv(here("data", "time_spent_in_lockdown_income.csv"), fileEncoding = "latin1")

# Remove unwanted columns (the things not needed for the graph)
df1 <- df1_orig %>%
  select(-X.3, -X.4, -X.5,
        -X.8, -X.9, -X.10,
        -X.11, -X.12, -X2014.2015.vs.March.April.2020)

# Make column names meaningful
names(df1)[names(df1) == "X"] <- "activity"
names(df1)[names(df1) == "X2014.2015"] <- "low_1415"
names(df1)[names(df1) == "X.1"] <- "med_1415"
names(df1)[names(df1) == "X.2"] <- "high_1415"
names(df1)[names(df1) == "March.April.2020"] <- "low_2020"
names(df1)[names(df1) == "X.6"] <- "med_2020"
names(df1)[names(df1) == "X.7"] <- "high_2020"

# Identify all cells that are empty as NA
df1[df1==""] <- NA

# Keep only rows with no NA
df1 <- na.omit(df1)

# Make the columns numeric
df1$low_1415 <- as.numeric(as.character(df1$low_1415))
df1$med_1415 <- as.numeric(as.character(df1$med_1415))
df1$high_1415 <- as.numeric(as.character(df1$high_1415))
df1$low_2020 <- as.numeric(as.character(df1$low_2020))
df1$med_2020 <- as.numeric(as.character(df1$med_2020))
df1$high_2020 <- as.numeric(as.character(df1$high_2020))

```

The first few rows of the newly organised data are displayed below.

```

##              activity low_1415 med_1415
## 3 Travelling and transport (e.g. walking or driving) 65.2    84.6
## 4              Working not from home 70.3   173.7
## 5              Working from home 6.4    11.4
## 6              Study 16.6    13.9
## 7              Keep fit 16.2    17.5
## 8              Unpaid childcare 28.2    36.1
##  high_1415 low_2020 med_2020 high_2020
## 3      102.1    21.6    16.9    17.1
## 4      202.0    65.7   102.1   138.1
## 5       24.0    32.4    51.1    89.3
## 6       13.6    11.7     5.7     5.9
## 7       23.3    16.4    23.6    31.7
## 8       30.9    33.6    40.3    35.5

```

The numeric values represent time (in minutes) spent doing each activity. Excluding the activity column, each column represents participants with either a high, medium or low income in the specified year.

Incomes were categorised based on the following income bands:

**Low income: up to £1,700 per month**

Medium income: £1,700 to £3,300 per month

High income: over £3,300 per month

The visualisation was improved by calculating the percentage change, as the relative difference between the two time periods was important, rather than their absolute differences. Specifically, this meant that the difference in time spent on activities, by people of different income levels between 2014/15 and 2020, was calculated, allowing a percentage difference to be determined.

The code for this process is presented below:

```
# Calculate percentage change columns
df2 <- df1 %>%
  mutate(low = ((low_2020 - low_1415)/low_1415)*100,
         med = ((med_2020 - med_1415)/med_1415)*100,
         high = ((high_2020 - high_1415)/high_1415)*100)
df2<- select(df2, low, med, high, activity)

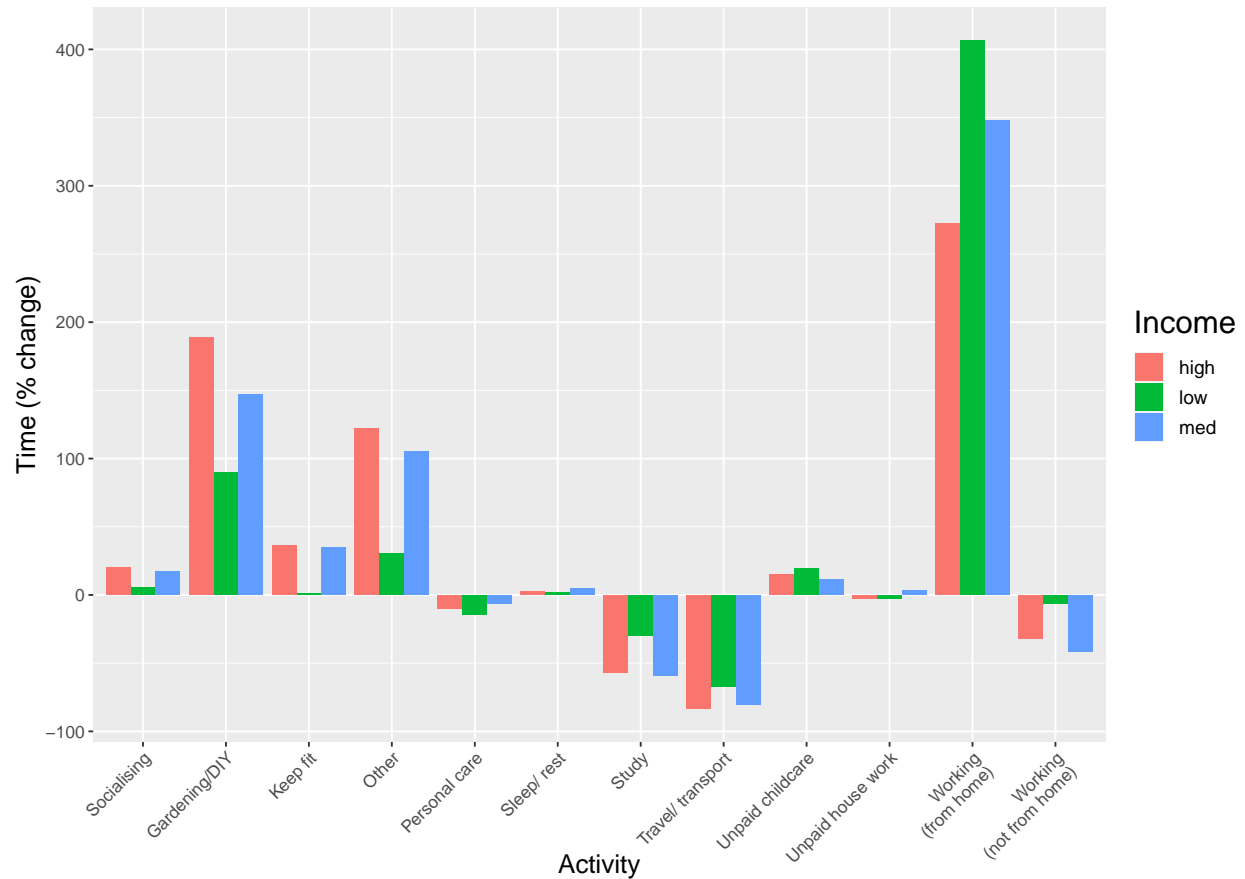
# Put percentage change columns into a long data frame
keycol <- "income"
valuecol <- "time_ch"
gathercols <- c("low", "med", "high")

df2 <- gather_(df2, keycol, valuecol, gathercols)

head(df2)

##               activity income    time_ch
## 1 Travelling and transport (e.g. walking or driving)    low -66.871166
## 2               Working not from home    low  -6.543385
## 3               Working from home    low 406.250000
## 4                   Study    low -29.518072
## 5                   Keep fit    low   1.234568
## 6             Unpaid childcare    low  19.148936
```

Below is a visualisation with all of the activities accounted for:



After plotting this graph, it was clear that some of the activities showed a negligible percentage change. The decision was made to exclude activities that had a low percentage change. To do this, the mean percentage change was calculated across the three income levels for each activity.

If the mean value for an activity (across the income levels) was between -16% and +16%, the data for that activity was excluded. To achieve this, as some values were negative, the square of the mean percentage change was calculated and compared with 256 ( $16^2$ ). The condition labeled “Other” was also excluded as it is uninformative.

This left the 6 activities that showed greater percentage change overall.

The code for this exclusion process is displayed below:

```
# Calculate mean by activity
df2 <- df2 %>%
  group_by(activity) %>%
  mutate(mean_by_activity = mean(time_ch))

# Put columns to 0 dp
df2 <- df2 %>% mutate_if(is.numeric, round, digits=0)

# Get rid of all values between -16 and 16.
```

```

keep_rows <- c()

for(i in 1:length(df2$mean_by_activity)) {
  if((df2$mean_by_activity[i])**2 > 256){
    keep_rows <- c(keep_rows, i)
    print(i)
  }
}

# Create new data frame that has only the rows with the greater percentage change
# Remove the "other" condition as it is uninformative
df3 <- df2[keep_rows,]
df3 <- df3[df3$activity != "Other",]

```

## The final visualisation

```

# Put the income levels in order and re-label them for the sake of the graph legend

df3$income <- factor(df3$income,
                     levels = c("high", "med", "low"),
                     labels = c("High", "Medium", "Low"))

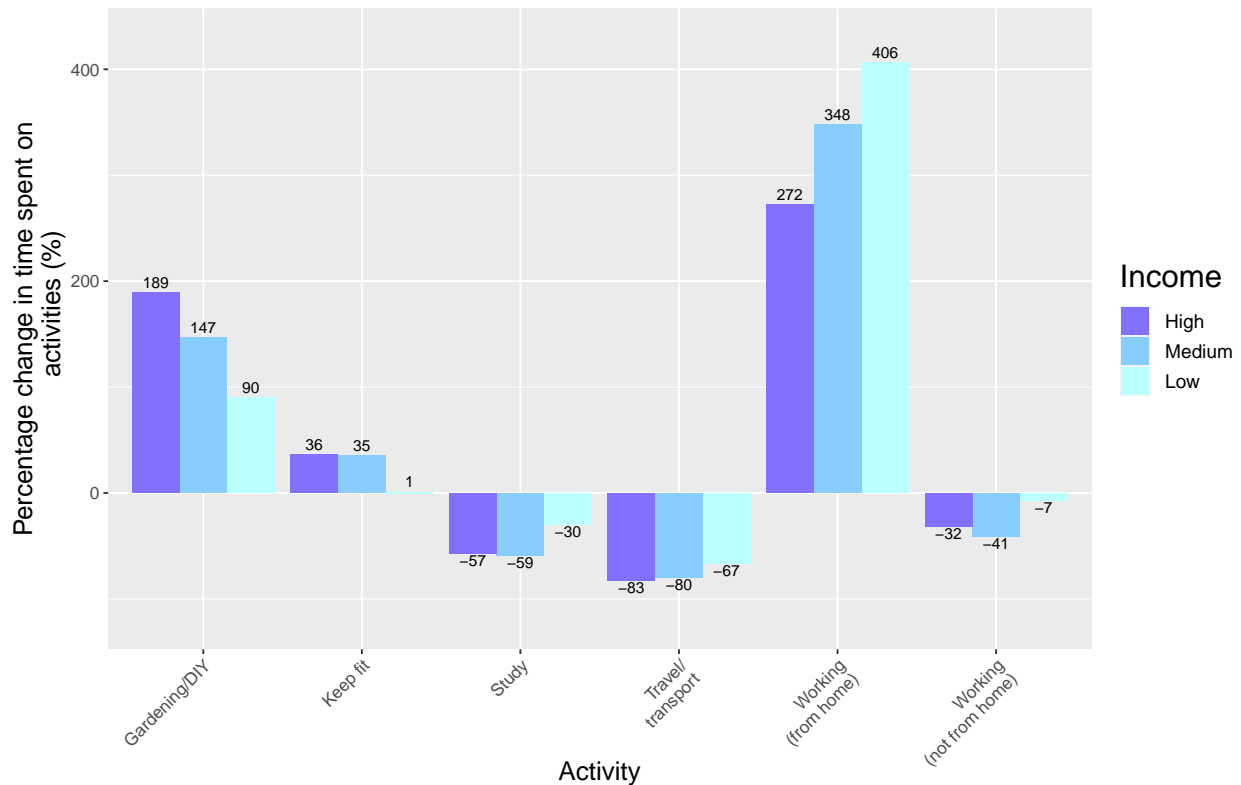
# Plot the graph
ggplot(df3, aes(y = time_ch, x = activity, fill = income)) +
  geom_col(position = position_dodge()) +
  geom_text(aes(label = paste(time_ch),
                             vjust = ifelse(time_ch >= 0, -0.4, 1.1)),
            position = position_dodge(0.9),
            size = 3) +
  scale_y_continuous(limits = c(-120, 430)) +
  scale_x_discrete(labels=c("Gardening/DIY",
                           "Keep fit",
                           "Study",
                           "Travel/
                           transport",
                           "Working
                           (from home)",
                           "Working
                           (not from home)")) +
  scale_fill_manual(values = c("High" = "lightslateblue",
                              "Medium" = "skyblue1",
                              "Low" = "paleturquoise1")) +
  theme(plot.title = element_text (size = 20, face = "bold"),
        plot.subtitle = element_text(face = "italic"),
        plot.caption = element_text(vjust = 30),
        axis.text.x = element_text(angle = 45, hjust=1, size = 10),
        axis.title.x = element_text(size = 15, vjust = 20),
        axis.title.y = element_text(size = 15),
        axis.text.y = element_text(size = 10),
        legend.title = element_text(size=18),
        legend.text = element_text(size=11)) +
  labs(y = "Percentage change in time spent on
        activities (%)",

```

```
x = "Activity",
fill = "Income",
title = "Behaviour change: 2014/15 compared to 2020",
subtitle = "A bar chart showing the percentage change in time spent doing specified activities, arranged by income.",
caption = "Source: Office of National Statistics")
```

## Behaviour change: 2014/15 compared to 2020

A bar chart showing the percentage change in time spent doing specified activities, arranged by income.



Source: Office of National Statistics

### Summary

Overall, less time was spent studying, travelling and working away from home. More time was spent doing gardening/ DIY, keeping fit, and working from home. The extent of the behavior change was different for people of different income brackets.

If there was more time available, other variables that are available from *the dataset* could be incorporated into the visualisation. In particular, it might be interesting to investigate whether participants that had children were affected similarly to those without children, both within and between their income brackets.

### Helpful resources:

Healy, K. (2018). *Data visualisation: a practical introduction*. Princeton University Press.