My local authority recently decided to publish all spending over £500 in an effort to be more transparent. Here’s a post taking an overview of what they’ve published. I’ve used R for the analysis. The dataset doesn’t contain much detail, but if you’ve analysis suggestions, please add them in the comments!

You can [download the spending data here](https://www.midlothian.gov.uk/downloads/download/672/expenditure_over_500). It’s available in pdf (why?!) and xlsx ([plain text would be more open](https://en.wikipedia.org/wiki/FAIR_data)).

First off, some packages:

library(tidyverse)

library(readxl)

library(janitor)

library(lubridate)

library(formattable)

Read in the dataset:

df = read\_excel("~/Downloads/midlothian\_payments\_over\_500\_01042019\_to\_15092019.xlsx") %>%

clean\_names()

We’ve got six columns:

* type
* date\_paid
* supplier
* amount
* our\_ref
* financial\_year

Busiest day:

df %>%

mutate(day = weekdays(date\_paid)) %>%

group\_by(day) %>%

summarise(transactions = n(),

thousands\_pounds\_spent = sum(amount) / 1000) %>%

mutate(day = fct\_relevel(day, rev(c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")))) %>%

gather(var, value, -day) %>%

ggplot(aes(day, value)) +

geom\_col() +

facet\_wrap(~var, scales = "free\_x") +

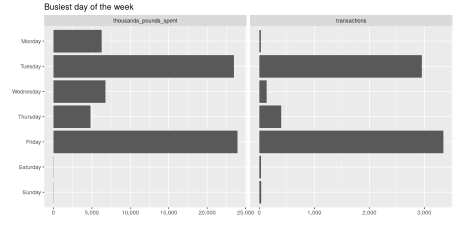
coord\_flip() +

scale\_y\_continuous(labels = scales::comma) +

labs(title = "Busiest day of the week",

x = "",

y = "")



Busiest time of year:

df %>%

mutate(dow = weekdays(date\_paid),

dow = if\_else(dow == "Tuesday" | dow == "Friday", "Tue/Fri", "Other")) %>%

group\_by(date\_paid, dow) %>%

summarise(transactions = n(),

pounds\_spent = sum(amount)) %>%

gather(var, value, -date\_paid, -dow) %>%

ggplot(aes(date\_paid, value, colour = dow)) +

geom\_point() +

facet\_wrap(~var, scales = "free\_y") +

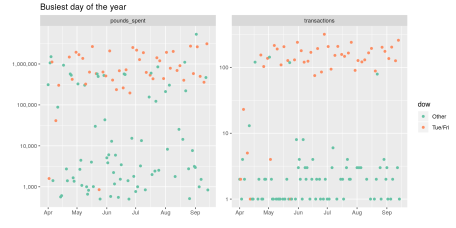
scale\_y\_log10(labels = scales::comma) +

scale\_colour\_brewer(type = "qual", palette = "Set2") +

labs(title = "Busiest day of the year",

x = "",

y = "")



Top 10 payees by value:

df %>%

group\_by(supplier) %>%

summarise(pounds\_spent = sum(amount),

transactions = n()) %>%

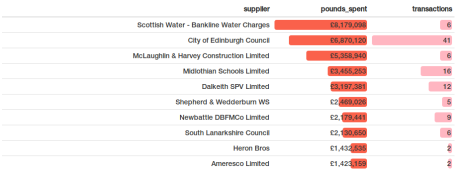
arrange(desc(pounds\_spent)) %>%

top\_n(n = 10, wt = pounds\_spent) %>%

mutate(pounds\_spent = currency(pounds\_spent, "£", digits = 0L)) %>%

formattable(list(`pounds\_spent` = color\_bar("#FA614B"),

`transactions` = color\_bar("lightpink")))



In Scotland local authorities collect water charges on behalf of the water authority, which they then pass on. It’s not surprise that Scottish Water is the biggest supplier.

Top 10 payees by frequency:

df %>%

group\_by(supplier) %>%

summarise(pounds\_spent = sum(amount),

transactions = n()) %>%

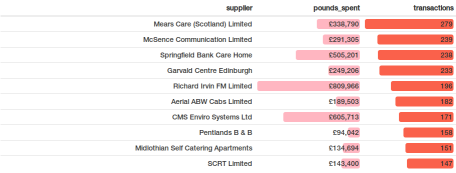
arrange(desc(transactions)) %>%

top\_n(n = 10, wt = transactions) %>%

mutate(pounds\_spent = currency(pounds\_spent, "£", digits = 0L)) %>%

formattable(list(`pounds\_spent` = color\_bar("lightpink"),

`transactions` = color\_bar("#FA614B")))



As a final note, writing this post is reminding me *again* I should be moving away from wordpress because incorporating code and output would be much easier with mark/blog down! As always, legacy is holding me back.