SFCC\_Internship\_Analyses

Hamish Williams

2024-07-04

# Working Notes Document - Week 2 SFCC Internship

In this document I will go through chronologically the tasks completed and discuss the results of the work completed to date. This markdown specifically is to serve as a reproducible and good example of how a markdown document can be made to inform others on the work completed with accessible coding notations, and in the instance of this internship and the goals set out at the start of the internship, to generate a reproducible document which could be used to further assess data in the future.

There will be several parts to this markdown. I will include spaces for the necessary packages and custom functions used to generate the results, as well as directing the user to the appropriate spaces where the code for the shiny apps produced are stored (note: I will include the coding for the apps in here, but will likely have to comment these coding sections to render them nonfunctional, otherwise the generation of this file will not be possible). I will also give some notes on my thoughts and ideas from the interpretations of the results, and how these could be utilized by the SFCC and FMS in the future to better collate and analyse their costings/shortfalls for the fisheries boards/trusts, whilst considering the issues and costs of altering the data gathering processes.

## GitHub

All of the work completed here is accesible from the GitHub page which I have created to work on the data given for the internship: (<https://github.com/HamishGWilliams/SFCC_Internship.git>). For the sake of transparency and accessibility to the GitHub, I have made this public currently so that users may access the github and view the data, results and apps.

## Shiny Apps

Speaking of apps, there are currently 2 live, which can be accessed from here: The first is the initial shiny app made which allows an interactive session to play around with a boxplot figure representing the funding shortfalls variable from the raw data (<https://hamish-williams.shinyapps.io/SFCC_Internship_Interactable_App_1/>), the second app is more refined and has had serveral to a dozen processes completed to generate the necessary results to form the plots; This second shiny app provides and interactble session where the user and select between the differenct actions, scales and set a costings threshold on a sliding scale to show the boxplots of the costings of the various actions across all districts (<https://hamish-williams.shinyapps.io/SFCC_Internship_Interactable_App_2/>).

More information on how these were created and the modifications to the data completed to generate the results for these apps will be shown in a later part of this document

# Setup: loading packages and writing custom functions

The following are the neccessary packages required to complete the analyses proceeding, if you do not have these packages currently downloaded, then use the commented example to download the package(s) required on your device:

# Example of package installation  
 # install.packages("ggplot2")  
  
# Load the packages  
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(ggplot2)  
library(shiny)  
library(here)

## Warning: package 'here' was built under R version 4.4.1

## here() starts at C:/Users/hamis/OneDrive/Documents/PhD/Internship/SFCC\_Internship

library(rsconnect)

##   
## Attaching package: 'rsconnect'

## The following object is masked from 'package:shiny':  
##   
## serverInfo

# Function to calculate means for a given scale  
calculate\_means <- function(data, scale) {  
 data %>%  
 group\_by(Proposed.Management.Action) %>%  
 summarize(average\_cost = mean(Filtered\_Costings\_Data, na.rm = TRUE)) %>%  
 mutate(scale = scale)  
}

# Excel work

A key issue exists with the raw excel data, which is that where and what the funding shortfall is based on is ambiguous. At times it seems to reflect the captial costings, the in-kind contributions, the annual costs, or even all 3 combined, whereas at othertimes it exists externally to the costings, and is what I would call a “True” shortfall; Which is to say that the shortfall given is not tied directly to the costings variables.

To compensate for this, an idea was proposed and reviewed by Sean, which is to devise a method to determine the likelihood that the funding shortfall is representative of the costings or not, and based on this combine all of the costings and shortfalls together to make a representative value for the total cost of the action.

The overall objective was to create a new data variable which instead of representing the funding shortfall, would represent a best estimate costs of the action of that row of data. Using this true cost, you could then subtract the funding shortfall from costings of this

There was a substantial amount of work completed not on this document, but on the excel spreadsheet (copy) of the data. I will take you through the steps completed to the excel data

## Key Considerations

From an email discussion with Sean, the following points were considered, quote:

1. If its 0 I think they will not need funding – there is another “Funding\_Status” column, where this is listed as “unfunded” and the shortfall is 0 then I would interpret that as the action has not been costed for whatever reason. If the funding status is “funded” or “partially funded” and the shortfall is 0, I would interpret that as not needing extra cash.
2. Yes, for those type of actions you have mentioned I think it would be appropriate to combine the capital/in-kind and annual costs to replace the 0 value in the “Funding Shortfall” column, as I know that is how some of the trusts calculated the shortfall. If a new field is created combining the shortfall and other costs then we may risk double-counting.

From this I have made the following assumptions:

* If the funding shortfall is unfunded AND = 0, then don’t include the shortfall in the calculation as it has not been costed for
* If the funding shortfall is unfunded BUT > 0, then include this in the calculations
* If the shortfall is funded OR > 0, then the shortfall could be included in the calculation
* IF the funding shortfall is not a reflection of either the capital, in-kind or annual costs (i.e., shortfall = cost), then the shortfall should be included in the calculation
* IF the funding shortfall is unfunded OR = 0, AND IS A reflection of any of the costings, then do not include the shortfall in the calculations

## Steps taken to generate new variable

The following steps were then followed to complete this set of criteria:

1. Make a new sheet with a copy of all the data
2. remove redundant variables not used in data, making sure to include the ID column in case there is need for one of the removed columns in the future
3. New Column: “Funding\_Status\_TRUE” - identifies whether the action has funding or not.

* ***=IF(OR($M2=“Partially funded”, $M2=“Funding secured”), “Funded”, “Unfunded”)***
* where M column = “Funding Available
* Key aim is to help discern if the funding shortfall is based on the costings variables or not, and was an advisement of Sean when asking for considerations for this objective

1. New column: ‘Shortfall\_Capital’ - identifies if shortfall reflects capital costings

* ***=IF(AND(Q2>0,Q2=I2),“Yes”,“No”)***
* Where Q = shortfall and I = capital costs

1. New columns: “Shortfall\_In-Kind” & “Shortfall\_Annual” - same as before, but for the other costings

* ***=IF(AND(Q2>0,Q2=J2),“Yes”,“No”)*** for the in-kind
* ***=IF(AND(Q2>0,Q2=L2),“Yes”,“No”)*** for the annual

1. New column: “Condition\_total\_costings” - uses the previously generated criteria to determine how if it calculates the costs including the shortfall or not:

* ***=IF(AND(OR(R2=“Funded”,Q2>0),S2=“No”,T2=“No”,U2=“No”),SUM(I2,J2,L2,Q2),““)***
* IF the criteria in the row shows that the action is funded OR has a shortfall > 0, AND the shortfall does NOT reflect any of the costings, then it will sum the values of all the costings and the shortfall. IF NOT, then it will leave a blank cell
* For brevities sake leaving a blank cell could be replaced with simply just calculating the costings together, but I did not do that before as I was getting confused with the excel coding. I instead made the combined costings in a new column

1. New column: “Combined\_Costings” - combines all the costing together
2. New column: “Chossing the costing variable” - selects which calculation to take show

* \*\*\*=IF(V2=““,W2,V2)\*\*\* IF the Condition variable =”“, then select the combined costings, otherwise use the conditioned variable.
* Again, would not need this if I had tidied up the coding more, but its not much more space needed to complete the step, and separating these steps also means that Its easier to make changes to individual steps at a time.

1. New column: “Filtered\_Costings\_Data” - removes any 0’s and replaces them with “” (NAs) as to not have these values skew the costings averages towards lower estimates.

* \*\*\*=IF(X2=0,““,X2)\*\*\* simply takes the value, IF = 0, then replace with””

I then saved this new sheet as “Fishery\_Management\_Plan\_Actions\_Working\_Sheet.csv”, which is then used in the analyses completed later.

This is not a perfect strategy, however I feel like it does a much better job to try and represent the actual costs of actions, but just the shortfalls. Using this new variable, we can then take the costings or the funding shortfalls and be able to calculate the aboslute and relative parts of the actions which are not accounted for with funding currently.

# R Work

The next step was to try and create some summary data of the costings, and to illustrate the distributions of the costings in a box plot

# Import Data  
data <- read.csv("../data/Fishery\_Management\_Plan\_Actions\_Working\_Sheet.csv", header = T)  
  
# Change data type of relevant variables  
str(data)

## 'data.frame': 1323 obs. of 25 variables:  
## $ ObjectID : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ Proposed.Management.Action : chr "D" "S" "M" "N" ...  
## $ Strategy : chr "Local engagement with local funders and landowners, Link landowners with trees for life project officer" "Liase with SSE and SEPA when necessary" "Local engagement with local landowners" "Board policy and staff engagemnet " ...  
## $ Scale : int 4 4 4 4 4 4 5 1 1 1 ...  
## $ Resources : chr "https://drive.google.com/file/d/1KujlrX9hbjv-hqaKBqnt97w51\_BF0gTb/view?usp=share\_link" "https://drive.google.com/file/d/1hheqTtZVABHyycjym7lCDCruDuNHDKnX/view?usp=share\_link" "" "https://docs.google.com/document/d/17h4fadHwVA\_k169BY7a6oxD6jgZY96KT/edit?usp=share\_link&ouid=10820169571111140"| \_\_truncated\_\_ ...  
## $ Data.Requirements : chr "Scottish River Temperature Monitoring Network maps on NMPi" "" "NatureScot GIS layers" "" ...  
## $ Lead.Organisation : chr "Beauly Fishery Board" "Beauly Fishery Board" "Beauly fishery board" "Beauly fishery board " ...  
## $ Potential.Partners : chr "Trees for Life, SSEN" "SSE, SEPA, FMS" "Landowners, NatureScot Peatland Action" "" ...  
## $ Capital.Cost : num 0 0 0 0 0 7500 0 30000 45000 45000 ...  
## $ In.Kind.Contributions : num 1400 1400 1400 0 700 700 41000 13000 15000 10000 ...  
## $ Maintainance.Required : chr "No" "No" "No" "No" ...  
## $ Annual.Cost : int 0 0 0 0 0 0 41000 13000 NA 0 ...  
## $ Funding.Available : chr "Unfunded" "Unfunded" "Unfunded" "Unfunded" ...  
## $ Completion.Date : chr "Within 2 Years" "Annual" "More than 5 Years" "Within 3 Years" ...  
## $ District : chr "Beauly" "Beauly" "Beauly" "Beauly" ...  
## $ Costing.Certainty : chr "" "" "" "" ...  
## $ Funding.Shortfall : int NA NA NA NA NA NA NA NA NA NA ...  
## $ Funding\_Status\_TRUE : chr "Unfunded" "Unfunded" "Unfunded" "Unfunded" ...  
## $ Shortfall\_Captial : chr "No" "No" "No" "No" ...  
## $ Shortfall\_In.Kind : chr "No" "No" "No" "No" ...  
## $ Shortfall\_Annual : chr "No" "No" "No" "No" ...  
## $ Conditioned\_Total\_Costings : num NA NA NA NA NA NA 82000 56000 NA NA ...  
## $ Combined\_Costings : num 1400 1400 1400 0 700 8200 82000 56000 60000 55000 ...  
## $ Choosing.the.Costing.Variable: num 1400 1400 1400 0 700 8200 82000 56000 60000 55000 ...  
## $ Filtered\_Costings\_Data : num 1400 1400 1400 NA 700 8200 82000 56000 60000 55000 ...

data$Proposed.Management.Action <- as.factor(data$Proposed.Management.Action)  
data$Scale <- as.factor(data$Scale)  
data$Proposed.Management.Action <- as.factor(data$Proposed.Management.Action)  
data$Maintainance.Required <- as.factor(data$Maintainance.Required)  
data$Funding.Available <- as.factor(data$Funding.Available)  
data$Funding\_Status\_TRUE <- as.factor(data$Funding\_Status\_TRUE)  
data$Maintainance.Required <- as.factor(data$Maintainance.Required)  
data$Costing.Certainty <- as.factor(data$Costing.Certainty)

# Subset to very confident funding shortfall data  
data\_subset <- data[data$Costing.Certainty == "Very Confident",]  
# Save this subsetted dataframe to a new file for use.  
write.csv(data\_subset, file = "../results/SFCC\_VC\_Subset.csv")  
  
# subset data to relevant scales for investigation: 1-3  
data\_scale\_1 <- data\_subset[data\_subset$Scale == "1",]  
str(data\_scale\_1)

## 'data.frame': 107 obs. of 25 variables:  
## $ ObjectID : int 2053 2056 2069 2073 2078 2080 2103 2104 2105 2154 ...  
## $ Proposed.Management.Action : Factor w/ 22 levels "","A","B","C",..: 3 7 7 3 3 11 19 7 21 15 ...  
## $ Strategy : chr "Assess size of boulders and employ suitable strategy to clear the pass entrance" "Long term disruption to sediment transport due to Orrin dam has left part of the lower Orrin devoid of suitable"| \_\_truncated\_\_ "Working with the local estate, introduce rootplates and salmon carcasses in a small reach of the Rannoch Burn" "To install and operate a Whoosh Innovations electronic guidance curtain to assist smolt navigation through Loch"| \_\_truncated\_\_ ...  
## $ Scale : Factor w/ 5 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Resources : chr "" "" "" "" ...  
## $ Data.Requirements : chr "spawning success and juvenile recruitment upstream of the fish pass" "Hydromorphology report, flow regime, freshet allowance" "Baseline juvenile sizes and densities. Location of stocking sites. Eggs put down" "Natural' survival rate through reserviors. " ...  
## $ Lead.Organisation : chr "CFDSB" "CFDSB" "Local DSFB" "" ...  
## $ Potential.Partners : chr "Strathconon Estate" "SSE, SEPA" "Strathvaich Estate" "SSE, Strathconon Estate" ...  
## $ Capital.Cost : num 1000 40000 1500 45000 50000 2000 1000 2000 2000 0 ...  
## $ In.Kind.Contributions : num 0 0 0 0 0 ...  
## $ Maintainance.Required : Factor w/ 3 levels "","No","Yes": 3 3 2 2 2 2 3 3 3 2 ...  
## $ Annual.Cost : int 250 400 0 3000 0 100 0 0 0 2000 ...  
## $ Funding.Available : Factor w/ 4 levels "","Funding Secured",..: 4 1 4 2 1 1 2 2 2 3 ...  
## $ Completion.Date : chr "Within 1 Year" "Within 2 Years" "Within 5 Years" "Within 3 Years" ...  
## $ District : chr "Cromarty" "Cromarty" "Cromarty" "Cromarty" ...  
## $ Costing.Certainty : Factor w/ 6 levels "","Estimate",..: 6 6 6 6 6 6 6 6 6 6 ...  
## $ Funding.Shortfall : int 0 0 0 0 50000 0 0 0 0 5000 ...  
## $ Funding\_Status\_TRUE : Factor w/ 2 levels "Funded","Unfunded": 2 2 2 1 2 2 1 1 1 1 ...  
## $ Shortfall\_Captial : chr "No" "No" "No" "No" ...  
## $ Shortfall\_In.Kind : chr "No" "No" "No" "No" ...  
## $ Shortfall\_Annual : chr "No" "No" "No" "No" ...  
## $ Conditioned\_Total\_Costings : num NA NA NA 48000 NA ...  
## $ Combined\_Costings : num 1250 40400 1500 48000 50000 ...  
## $ Choosing.the.Costing.Variable: num 1250 40400 1500 48000 50000 ...  
## $ Filtered\_Costings\_Data : num 1250 40400 1500 48000 50000 ...

data\_scale\_2 <- data\_subset[data\_subset$Scale == "2",]  
str(data\_scale\_2)

## 'data.frame': 7 obs. of 25 variables:  
## $ ObjectID : int 2085 2091 2279 2301 2910 2970 3145  
## $ Proposed.Management.Action : Factor w/ 22 levels "","A","B","C",..: 9 11 18 6 6 5 9  
## $ Strategy : chr "Response informed by current science and best magement practice. Taking advice from MSS and ture Scot. The Boar"| \_\_truncated\_\_ "Seek for research to be carried out into the impact of dolphins on Dee salmon stocks in the harbour. Healthy sa"| \_\_truncated\_\_ "SLRT and ZSL tagged sea trout from 2 adjacent sea loch systems in 2021 and 2022 and tracked their movements in "| \_\_truncated\_\_ "1) Work with Marine Scotland to conduct regular gillnetting searches in the coastal zone 2) Investigate reports"| \_\_truncated\_\_ ...  
## $ Scale : Factor w/ 5 levels "1","2","3","4",..: 2 2 2 2 2 2 2  
## $ Resources : chr "Vattenfall report, Previous harbour smolt tracking reports" "" "https://slrt.org.uk/juvenile-salmonid-surveys-3/" "" ...  
## $ Data.Requirements : chr "A better understanding of the migration routes of smolts and adult salmonids" "Accurate understanding of dolphin presence and behaviour. Also, better knowledge of dolphin diet from this area." "" "" ...  
## $ Lead.Organisation : chr "Dee DSFB" "Dee DSFB" "ZSL" "SFB" ...  
## $ Potential.Partners : chr "RDT, ture Scot, SEPA, FMS, MSS, MS LOT, Port of Aberdeen, Developers" "SMRU, FMS, MSS, tureScot, ShoreWatch" "SLRT, local ghillie" "" ...  
## $ Capital.Cost : num 0 0 250000 0 0 70000 0  
## $ In.Kind.Contributions : num 4000 950 0 9000 0 0 800  
## $ Maintainance.Required : Factor w/ 3 levels "","No","Yes": 2 2 2 3 2 2 2  
## $ Annual.Cost : int 0 0 250000 0 713 0 0  
## $ Funding.Available : Factor w/ 4 levels "","Funding Secured",..: 4 4 2 4 2 2 4  
## $ Completion.Date : chr "Annual" "Annual" "Within 2 Years" "Annual" ...  
## $ District : chr "Dee" "Dee" "Skye" "Spey" ...  
## $ Costing.Certainty : Factor w/ 6 levels "","Estimate",..: 6 6 6 6 6 6 6  
## $ Funding.Shortfall : int 0 0 0 9000 0 0 0  
## $ Funding\_Status\_TRUE : Factor w/ 2 levels "Funded","Unfunded": 2 2 1 2 1 1 2  
## $ Shortfall\_Captial : chr "No" "No" "No" "No" ...  
## $ Shortfall\_In.Kind : chr "No" "No" "No" "Yes" ...  
## $ Shortfall\_Annual : chr "No" "No" "No" "No" ...  
## $ Conditioned\_Total\_Costings : num NA NA 500000 NA 713 70000 NA  
## $ Combined\_Costings : num 4000 950 500000 9000 713 70000 800  
## $ Choosing.the.Costing.Variable: num 4000 950 500000 9000 713 70000 800  
## $ Filtered\_Costings\_Data : num 4000 950 500000 9000 713 70000 800

data\_scale\_3 <- data\_subset[data\_subset$Scale == "3",]  
str(data\_scale\_3)

## 'data.frame': 52 obs. of 25 variables:  
## $ ObjectID : int 2050 2065 2077 2082 2083 2084 2089 2094 2095 2145 ...  
## $ Proposed.Management.Action : Factor w/ 22 levels "","A","B","C",..: 17 3 19 13 13 13 11 13 16 18 ...  
## $ Strategy : chr "Sample of Meig 2023 smolt run to determine the number of breeders contributing to the run above Loch Meig. All "| \_\_truncated\_\_ "Working with SSE to continue hatchery operations" "Plant 1000 eggs per box at desired stocking density in the Blackwater mainstem. Egg boxes to be placed in reach"| \_\_truncated\_\_ "Undertake electrofishing and inverterbrate and nutrient surveys across a panel of index sites" ...  
## $ Scale : Factor w/ 5 levels "1","2","3","4",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ Resources : chr "" "" "" "" ...  
## $ Data.Requirements : chr "Genetic samples" "" "Stocking density requirements, juvenile density data" "" ...  
## $ Lead.Organisation : chr "CFDSB" "Local DSFB" "Local DSFB" "Local DSFB" ...  
## $ Potential.Partners : chr "Cromarty Firth Fisheries Trust, Institute of Biology and Freshwater conservation" "SSE" "SSE" "Strathconon estate" ...  
## $ Capital.Cost : num 25000 0 1000 0 0 0 1000 0 3500 0 ...  
## $ In.Kind.Contributions : num 0 0 0 0 0 ...  
## $ Maintainance.Required : Factor w/ 3 levels "","No","Yes": 3 1 2 1 1 1 2 2 2 2 ...  
## $ Annual.Cost : int 10000 182314 1000 5000 5000 5000 13646 0 2500 0 ...  
## $ Funding.Available : Factor w/ 4 levels "","Funding Secured",..: 2 2 2 4 4 4 4 4 4 4 ...  
## $ Completion.Date : chr "Within 5 Years" "Annual" "N/A" "More than 5 Years" ...  
## $ District : chr "Cromarty" "Cromarty" "Cromarty" "Cromarty" ...  
## $ Costing.Certainty : Factor w/ 6 levels "","Estimate",..: 6 6 6 6 6 6 6 6 6 6 ...  
## $ Funding.Shortfall : int 0 0 0 0 0 0 4500 0 5000 0 ...  
## $ Funding\_Status\_TRUE : Factor w/ 2 levels "Funded","Unfunded": 1 1 1 2 2 2 2 2 2 2 ...  
## $ Shortfall\_Captial : chr "No" "No" "No" "No" ...  
## $ Shortfall\_In.Kind : chr "No" "No" "No" "No" ...  
## $ Shortfall\_Annual : chr "No" "No" "No" "No" ...  
## $ Conditioned\_Total\_Costings : num 35000 182314 2000 NA NA ...  
## $ Combined\_Costings : num 35000 182314 2000 5000 5000 ...  
## $ Choosing.the.Costing.Variable: num 35000 182314 2000 5000 5000 ...  
## $ Filtered\_Costings\_Data : num 35000 182314 2000 5000 5000 ...

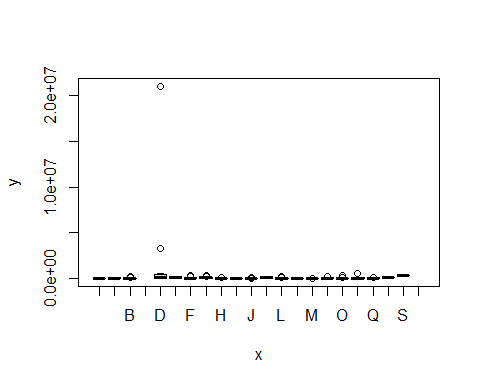
data\_scale\_4 <- data\_subset[data\_subset$Scale == "4",]  
str(data\_scale\_4)

## 'data.frame': 72 obs. of 25 variables:  
## $ ObjectID : int 1936 1959 2071 2159 2178 2181 2184 2230 2260 2261 ...  
## $ Proposed.Management.Action : Factor w/ 22 levels "","A","B","C",..: 10 13 3 10 17 10 9 13 17 17 ...  
## $ Strategy : chr "Lobby for inclusion of riparian planting as a priority for future iterations of WGS." "Continual monitoring, develop plan for survey work and structure." "Capture of smolts in wolf and rotary screw traps, transport downstream of major hydro in aerated water transport tanks" "Bladnoch DSFB/GFT to campaign publicly their opposition to replanting conifers on deep peat and why - especiall"| \_\_truncated\_\_ ...  
## $ Scale : Factor w/ 5 levels "1","2","3","4",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ Resources : chr "" "DSFB" "" "" ...  
## $ Data.Requirements : chr "" "" "numbers of adults returning from sea. Estimates of sea survival" "None" ...  
## $ Lead.Organisation : chr "Landowner with assistance from the DSFB" "DSFB" "DSFB" "Bladnoch DSFB" ...  
## $ Potential.Partners : chr "FMS, Woodland Trust, etc." "DSFB" "SSE, Strathconon Estate" "GFT, Crichton Carbon Centre" ...  
## $ Capital.Cost : num 0 0 10000 0 0 0 0 0 15000 1500 ...  
## $ In.Kind.Contributions : num 0 10000 0 0 1000 0 0 0 20000 20000 ...  
## $ Maintainance.Required : Factor w/ 3 levels "","No","Yes": 3 1 3 2 3 2 2 1 3 3 ...  
## $ Annual.Cost : int 0 0 10000 1000 5000 1000 1050 0 15000 21500 ...  
## $ Funding.Available : Factor w/ 4 levels "","Funding Secured",..: 4 4 2 2 2 2 2 4 3 2 ...  
## $ Completion.Date : chr "Within 3 Years" "Annual" "N/A" "Annual" ...  
## $ District : chr "Brora" "Brora" "Cromarty" "Bladnoch" ...  
## $ Costing.Certainty : Factor w/ 6 levels "","Estimate",..: 6 6 6 6 6 6 6 6 6 6 ...  
## $ Funding.Shortfall : int 0 0 0 0 20000 0 0 0 15000 0 ...  
## $ Funding\_Status\_TRUE : Factor w/ 2 levels "Funded","Unfunded": 2 2 1 1 1 1 1 2 1 1 ...  
## $ Shortfall\_Captial : chr "No" "No" "No" "No" ...  
## $ Shortfall\_In.Kind : chr "No" "No" "No" "No" ...  
## $ Shortfall\_Annual : chr "No" "No" "No" "No" ...  
## $ Conditioned\_Total\_Costings : num NA NA 20000 1000 26000 1000 1050 NA NA 43000 ...  
## $ Combined\_Costings : num 0 10000 20000 1000 6000 1000 1050 0 50000 43000 ...  
## $ Choosing.the.Costing.Variable: num 0 10000 20000 1000 26000 1000 1050 0 50000 43000 ...  
## $ Filtered\_Costings\_Data : num NA 10000 20000 1000 26000 1000 1050 NA 50000 43000 ...

data\_scale\_5 <- data\_subset[data\_subset$Scale == "5",]  
str(data\_scale\_5)

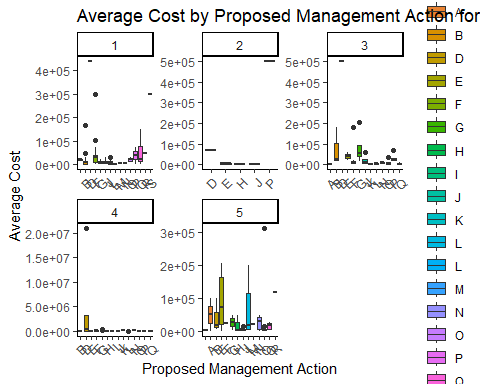
## 'data.frame': 77 obs. of 25 variables:  
## $ ObjectID : int 1960 1961 2044 2063 2064 2076 2081 2086 2087 2088 ...  
## $ Proposed.Management.Action : Factor w/ 22 levels "","A","B","C",..: 1 2 16 8 11 11 8 9 16 6 ...  
## $ Strategy : chr "BDSFB staff should receive ongoing training to maintain or increase skills for delivery of required tasks in re"| \_\_truncated\_\_ "Angling training days, free fly rod for under 18's to borrow for the season, free fishing permits and school trips." "Working with  the SISI project to engage with local communities and volunteers to treat and monitor INNS within the district" "Maintain effective bailiffing strategy and increase educational efforts across the district" ...  
## $ Scale : Factor w/ 5 levels "1","2","3","4",..: 5 5 5 5 5 5 5 5 5 5 ...  
## $ Resources : chr "DSFB and local community" "DSFB" "" "" ...  
## $ Data.Requirements : chr "" "" "Landownership, INNS distribution" "Internal bailiffing records" ...  
## $ Lead.Organisation : chr "DSFB" "DSFB & Local community" "SISI (Nature Scot)" "DSFB" ...  
## $ Potential.Partners : chr "DSFB" "DSFB & Local community" "" "Police Scotland" ...  
## $ Capital.Cost : num 0 0 15000 0 0 0 4000 0 1000 1000 ...  
## $ In.Kind.Contributions : num 2000 500 0 0 0 ...  
## $ Maintainance.Required : Factor w/ 3 levels "","No","Yes": 1 1 3 3 3 3 3 2 2 2 ...  
## $ Annual.Cost : int 0 0 15000 50000 6000 6000 200 0 0 80080 ...  
## $ Funding.Available : Factor w/ 4 levels "","Funding Secured",..: 2 2 2 2 2 2 4 4 4 4 ...  
## $ Completion.Date : chr "Annual" "Annual" "Annual" "Annual" ...  
## $ District : chr "Brora" "Brora" "Cromarty" "Cromarty" ...  
## $ Costing.Certainty : Factor w/ 6 levels "","Estimate",..: 6 6 6 6 6 6 6 6 6 6 ...  
## $ Funding.Shortfall : int 0 400 0 0 0 0 0 0 0 25000 ...  
## $ Funding\_Status\_TRUE : Factor w/ 2 levels "Funded","Unfunded": 1 1 1 1 1 1 2 2 2 2 ...  
## $ Shortfall\_Captial : chr "No" "No" "No" "No" ...  
## $ Shortfall\_In.Kind : chr "No" "No" "No" "No" ...  
## $ Shortfall\_Annual : chr "No" "No" "No" "No" ...  
## $ Conditioned\_Total\_Costings : num 2000 900 30000 50000 6000 ...  
## $ Combined\_Costings : num 2000 500 30000 50000 6000 ...  
## $ Choosing.the.Costing.Variable: num 2000 900 30000 50000 6000 ...  
## $ Filtered\_Costings\_Data : num 2000 900 30000 50000 6000 ...

# Boxplot of filtered costings data ~ actions  
plot(data\_subset$Proposed.Management.Action, data\_subset$Filtered\_Costings\_Data)



# ggplot  
ggplot(data\_subset, aes(x = Proposed.Management.Action, y = Filtered\_Costings\_Data, fill = Proposed.Management.Action)) +  
 geom\_boxplot() +   
 facet\_wrap(~Scale, scales = "free") +   
 labs(title = "Average Cost by Proposed Management Action for all Scales",  
 x = "Proposed Management Action",  
 y = "Average Cost") +  
 theme\_classic() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) +  
 scale\_fill\_discrete(name = "Action")

## Warning: Removed 12 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).



# Investigating the means of each scale for their costs per action group:  
by(data\_scale\_1$Filtered\_Costings\_Data, data\_scale\_1$Proposed.Management.Action, summary)

## data\_scale\_1$Proposed.Management.Action:   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 18400 20510 22620 22620 24730 26840   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: A  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: B  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 500 3000 10000 15012 14000 165000   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: C  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: D  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 439000 439000 439000 439000 439000 439000   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: E  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: F  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1100 8200 30640 50989 40400 297408   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: G  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2000 5000 7800 10520 15280 22520   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: H  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: I  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: J  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2100 4275 7500 11775 15000 30000   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: K  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: L  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1000 3500 6000 10221 13775 30000   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: L   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2000 2000 2000 2000 2000 2000   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: M  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 6000 6000 6500 6500 7000 7000   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: N  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 3900 3900 3900 3900 3900 3900   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: O  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 8000 14500 21000 19667 25500 30000   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: P  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2500 20625 38750 38750 56875 75000   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: Q  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2800 14880 22400 42800 77480 148704   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: R  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 50000 50000 50000 50000 50000 50000   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: S  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 297408 297408 297408 297408 297408 297408   
## ------------------------------------------------------------   
## data\_scale\_1$Proposed.Management.Action: T  
## NULL

by(data\_scale\_2$Filtered\_Costings\_Data, data\_scale\_2$Proposed.Management.Action, summary)

## data\_scale\_2$Proposed.Management.Action:   
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: A  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: B  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: C  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: D  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 70000 70000 70000 70000 70000 70000   
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: E  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 713 2785 4856 4856 6928 9000   
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: F  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: G  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: H  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 800 1600 2400 2400 3200 4000   
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: I  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: J  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 950 950 950 950 950 950   
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: K  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: L  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: L   
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: M  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: N  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: O  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: P  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5e+05 5e+05 5e+05 5e+05 5e+05 5e+05   
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: Q  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: R  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: S  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_2$Proposed.Management.Action: T  
## NULL

by(data\_scale\_3$Filtered\_Costings\_Data, data\_scale\_3$Proposed.Management.Action, summary)

## data\_scale\_3$Proposed.Management.Action:   
## NULL  
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: A  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2000 2000 2000 2000 2000 2000   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: B  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 15750 20375 25000 74355 103657 182314   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: C  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: D  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5e+05 5e+05 5e+05 5e+05 5e+05 5e+05   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: E  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 23758 32819 41879 41879 50940 60000   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: F  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5000 5500 10100 34457 17550 180000   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: G  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 16800 39900 53800 81800 95700 202800   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: H  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: I  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: J  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2000 5429 9920 19246 24756 60000   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: K  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 1000 1000 1000 1000 1000 1000 1   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: L  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1062 2155 5000 4507 5500 10000   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: L   
## NULL  
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: M  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: N  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 6312 7909 9506 9506 11103 12700   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: O  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1000 4000 5000 11500 12500 35000   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: P  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 19650 20000 20000 32330 32000 70000 1   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: Q  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2000 2000 2000 4667 6000 10000   
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: R  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: S  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_3$Proposed.Management.Action: T  
## NULL

by(data\_scale\_4$Filtered\_Costings\_Data, data\_scale\_4$Proposed.Management.Action, summary)

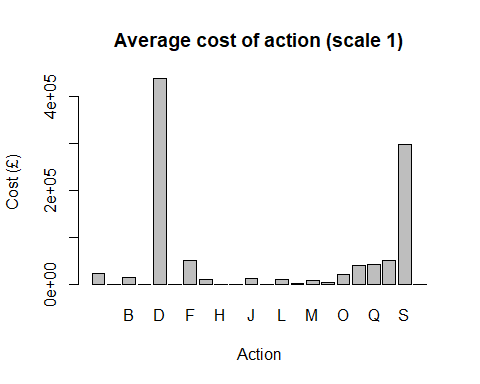
## data\_scale\_4$Proposed.Management.Action:   
## NULL  
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: A  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: B  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 2000 11000 20000 40667 60000 100000 1   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: C  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: D  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 6500 11000 335000 4926500 3230000 21050000   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: E  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1684 16263 30842 30842 45421 60000   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: F  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 50000 83900 122600 134650 173350 243400 1   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: G  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 10110 39760 69522 85895 89370 299360   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: H  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1050 2138 11250 13388 22500 30000   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: I  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 1000 1000 1000 1000 1000 1000 1   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: J  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4000 25405 46811 46811 68216 89622   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: K  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 160000 160000 160000 160000 160000 160000   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: L  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 7250 10500 13500 14708 15750 28000 2   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: L   
## NULL  
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: M  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: N  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 175000 193750 212500 212500 231250 250000   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: O  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5000 9000 30000 30663 42000 89622   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: P  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 81400 81400 81400 81400 81400 81400   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: Q  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2000 37660 73320 72440 107660 142000   
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: R  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: S  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_4$Proposed.Management.Action: T  
## NULL

by(data\_scale\_5$Filtered\_Costings\_Data, data\_scale\_5$Proposed.Management.Action, summary)

## data\_scale\_5$Proposed.Management.Action:   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2000 3000 4000 4000 5000 6000   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: A  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 900 25675 50450 50450 75225 100000   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: B  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: C  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: D  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5000 11000 17000 40667 58500 100000   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: E  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 350 22000 72290 91966 163370 207000 2   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: F  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 25000 25000 25000 25000 25000 25000   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: G  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4200 15650 27100 27100 38550 50000   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: H  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2000 3000 4000 18667 27000 50000   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: I  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: J  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 500 4500 6000 6611 7500 16000   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: K  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: L  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 3000 7500 18374 62690 115500 200000 3   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: L   
## NULL  
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: M  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 20000 20000 20000 20000 20000 20000   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: N  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 500 5000 30000 24864 43750 50000   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: O  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 920 3900 5000 39647 7000 314000   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: P  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: Q  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 3300 7200 19400 15995 23104 26400   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: R  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 120000 120000 120000 120000 120000 120000   
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: S  
## NULL  
## ------------------------------------------------------------   
## data\_scale\_5$Proposed.Management.Action: T  
## NULL

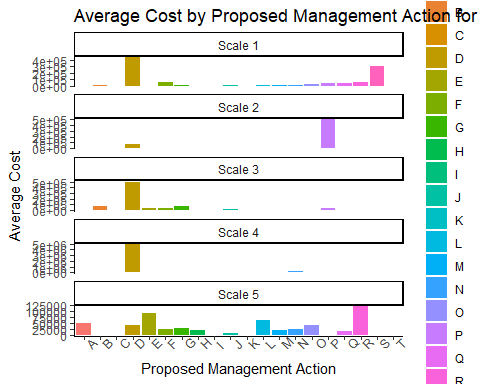
# Initial observations tell us that not all scales have points on all management actions.  
 # We can also see instances where there are very few datapoints for some actions at given scales  
 # another observation is that the mean and median values of actions or scales are not equal, which shows bias towards either the lower or upper values no the distrbution curves (assuming normall ditributed). This is likely not a concern, but something to consider, that some of the actions may have some biasing datapoints which pull the data in a particular direction  
 # Another pointto add to the last, the distributions are limited to a mimimum value of 0 (which in this instance has not been included as detailed in the excel section).  
  
# Function to calculate means for a given scale  
calculate\_means <- function(data, scale) {  
 data %>%  
 group\_by(Proposed.Management.Action) %>%  
 summarize(average\_cost = mean(Filtered\_Costings\_Data, na.rm = TRUE)) %>%  
 mutate(scale = scale)  
}  
  
# Calculate means for each scale  
results\_scale\_1 <- calculate\_means(data\_scale\_1, "Scale 1")  
results\_scale\_2 <- calculate\_means(data\_scale\_2, "Scale 2")  
results\_scale\_3 <- calculate\_means(data\_scale\_3, "Scale 3")  
results\_scale\_4 <- calculate\_means(data\_scale\_4, "Scale 4")  
results\_scale\_5 <- calculate\_means(data\_scale\_5, "Scale 5")  
  
# Combine results into a single data frame  
results <- bind\_rows(results\_scale\_1, results\_scale\_2, results\_scale\_3, results\_scale\_4, results\_scale\_5)  
  
# Ensure all actions (A:T) are represented for each scale  
all\_actions <- expand.grid(  
 scale = unique(results$scale),  
 Proposed.Management.Action = factor(LETTERS[1:20])  
)  
  
# Merge to ensure all actions are present, filling with NA where no data is available  
results\_full <- merge(all\_actions, results, by = c("scale", "Proposed.Management.Action"), all.x = TRUE)  
write.csv(results\_full, file = "../results/average\_spending\_results.csv")

# Barplot of average results  
barplot(results\_scale\_1$average\_cost~results\_scale\_1$Proposed.Management.Action,  
 xlab = "Action", ylab = "Cost (£)", main = "Average cost of action (scale 1)")



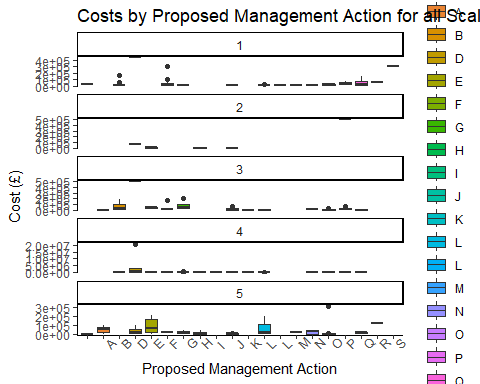
ggplot(results\_full, aes(x = Proposed.Management.Action, y = average\_cost, fill = Proposed.Management.Action)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Average Cost by Proposed Management Action for all Scales",  
 x = "Proposed Management Action",  
 y = "Average Cost") +  
 theme\_classic() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 0)) +  
 scale\_fill\_discrete(name = "Action") +   
 facet\_wrap(~scale, scales = "free\_y", nrow = 5)

## Warning: Removed 42 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).



# Boxplot of filtered costings data ~ actions  
ggplot(data\_subset, aes(x = Proposed.Management.Action, y = Filtered\_Costings\_Data, fill = Proposed.Management.Action)) +  
 geom\_boxplot() +   
 labs(title = "Costs by Proposed Management Action for all Scales",  
 x = "Proposed Management Action",  
 y = "Cost (£)") +  
 theme\_classic() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 0)) +  
 scale\_fill\_discrete(name = "Action") +   
 facet\_wrap(~Scale, scales = "free\_y", nrow = 5)

## Warning: Removed 12 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).



These plots are useful and provide a good basis to complete future work for refinement and presentation, and are submitted as such in this document.

Looking at the box plot, we can see, like mentioned in the comments of one of the last coding chunks, that there are some actions of different scales which are skewing the data quite a lot (represented here by outlier points on the boxplots). These limit our ability to see the shapes and distributions of the boxplots, however we should not throw these values to the way-side just yet, as they likely do represent real costs to specific actions.

What might be or more value to these plots is to further investigate the distribution of the costings across the districts and to see if these high values are all tied to a specific district, which might have certain characteristics and circumstances where these actions cost a great deal more for these specific actions on these districts.

Another observation is that there are a number of actions with little to no data points. Which is to say, that it is difficult to determine what their “likely average” costs are due to insufficient data present at the time. Given the fact that there is quite a range of costs for any given action (at each scale and also at each confidence - which is not shown at this time), which gives me cause for concern about trying to apply average values from these low sample data variables to provide “better” estimates for actions with lower estimate confidences.

That isn’t to say that these CANT be utilized, but if they are, they should be done with the consideration that there is possibilities that they may not be any more effective at estimate the total costs of a given action than professional opinion from contractors or inside personalle with experience in the managements, executions and therefore the costings of management actions.