Kat’s Leak Analysis by Water Provider & Text Analysis of Water Detection Type

Purpose

- The purpose of this file is to explore the HB1051 data with emphasis on water leakage. The data can be found http://cowaterefficiency.com/unauthenticated_home (http://cowaterefficiency.com/unauthenticated_home) with permission. Once in the portal, all report years (2013-2017) were selected as well as all water providers.
- Section 1.0 is clean up, section 2.10 is analysis.
- leak2.rmd & leak2.pdf - The purpose of this file is to explore the HB1051 data with emphasis on water leakage. Files were merged to created water leakage reporting by water provider and year. Data is displayed graphically and recommendations are provided for cleaning up potential reporting errors. The .rmd contains the raw markdown file, the pdf is just a printed version of this.
- ngrams.rmd & n.grams.pdf - The purpose of this file is to conduct text analysis on the leak_detection_field_technology column in the foundational_06a_loss_and_leak_detection file. Distributions of time between testing and age to replace small meters in years are shown. Recommendations for reporting are suggested to leverage this information. The .rmd file contains the raw markdown file, the pdf is just a printed version of this.

1.0 Reading in Data

- The following files were used:
  - overview.csv
  - normalizing_03_population.csv
  - foundational_09_balance_data.csv

```
#read in data
overview <- read.csv("EffDataPortal_Output_User690_20181112192716//overview/overview.csv", stringsAsFactors = FALSE)

normalizing_pop <- read.csv("EffDataPortal_Output_User690_20181112192716/normalizing/normalizing_03_population.csv", strings
AsFactors = FALSE)

foundational_09_balance_data <- read.csv("EffDataPortal_Output_User690_20181112192716/foundational/foundational_09_balance_d
ata.csv", stringsAsFactors = FALSE)

head(overview, n=10)
```

```
##    ce_annual_ndx ce_index                           water_provider
## 1           1936      760                       Westminster, City of
## 2           1944      762    Widefield Water and Sanitation District
## 3           1968      609                        Lafayette, City of
## 4           1990      663                       Northglenn, City of
## 5           1948      663                       Northglenn, City of
## 6           2007      573                   Grand Junction, City of
## 7           1913      499                          Cortez, City of
## 8           1967      623                        Longmont, City of
## 9           1926      647 Mount Werner Water and Sanitation District
## 10          1945      534         East Larimer County Water District
##    bndss_basin_index        basin bndss_type_index     type
## 1                  4        Metro                1  Utility
## 2                  1     Arkansas                5  Company
## 3                  7 South Platte                1  Utility
## 4                  7 South Platte                1  Utility
## 5                  7 South Platte                1  Utility
## 6                  3     Gunnison                1  Utility
## 7                  8    Southwest                1  Utility
## 8                  7 South Platte                1  Utility
## 9                 10  Yampa/White                2 Title 32
## 10                 7 South Platte                2 Title 32
##              county
## 1  Adams, Jefferson
## 2           El Paso
## 3           Boulder
## 4             Adams
## 5             Adams
## 6              Mesa
## 7         Montezuma
## 8           Boulder
## 9             Routt
## 10
##                                                                  provider_comments
## 1
## 2
## 3
## 4                  Water treatment in Adams County, wastewater treatment in Weld County
## 5                  Water treatment in Adams County, wastewater treatment in Weld County
## 6                                                          City of Grand Junction
```

```
## 7
## 8
## 9  MWWD operates raw water supply and treatment for MWWD and City of Steamboat Springs.
## 10
##    report_year report_status      bndss_id
## 1         2013     Submitted   Westminster
## 2         2013     Submitted  WidefieldWSD
## 3         2013     Submitted     Lafayette
## 4         2014     Submitted    Northglenn
## 5         2013     Submitted    Northglenn
## 6         2014     Submitted GrandJunction
## 7         2013     Submitted        Cortez
## 8         2013     Submitted      Longmont
## 9         2013     Submitted   MtWernerWSD
## 10        2013     Submitted          ELCO
```

```
head(normalizing_pop, n=10)
```

```
##    ce_annual_ndx pop_served
## 1           1936     117011
## 2           1944      17365
## 3           1968      26629
## 4           1990      35789
## 5           1948      35789
## 6           2007      27516
## 7           1913       8700
## 8           1967      93000
## 9           1926       4980
## 10          1945      19427
##                                                                                pop_served_source
## 1                                                 Official Westminster, Shaw Heights, out of City customers
## 2                                                              calculated by # sf & mf accts * 2.6 per acct
## 3                          Community Development Department data based on 2010 Census plus residential growth
## 4                                                                                             DRCOG
## 5                                                                                             DRCOG
## 6                                                                                         2010 Census
## 7                                                                                 CDPHE PWSID CO0142200
## 8                                                          Estimate for  inside City, outside taps, and Town of Lyons
## 9  City of Steamboat Springs data (4230 residents+ estimated 750 non-transients (employees not living in District)
## 10
##    transient_pop_served transient_pop_type_desc transient_pop_source
## 1                    NA      - None Specified -
## 2                     0      - None Specified -
## 3                     0      - None Specified -                None
## 4                     0      - None Specified -                none
## 5                    NA      - None Specified -                  na
## 6                     0      - None Specified -
## 7                    NA      - None Specified -
## 8                    NA      - None Specified -
## 9                 12000               Seasonal            estimate
## 10                   NA      - None Specified -
##                                                                remark
## 1
## 2
## 3                                                  Not significant
## 4                                                               na
## 5                                                               na
## 6
```

```
## 7
## 8
## 9  principally ski resort visitors Dec 20-Apr 10 and second homeowners
## 10
```

```
head(foundational_09_balance_data, n=10)
```

```
##    ce_annual_ndx water_type_index         water_type distributed_water
## 1           1912                1    Potable Treated       1.46120e+07
## 2           1912                2    Non-potable Raw       3.58762e+05
## 3           1912                3 Non-potable Re-use       3.44000e+05
## 4           1913                1    Potable Treated       6.45870e+08
## 5           1916                1    Potable Treated       4.36961e+02
## 6           1916                2    Non-potable Raw       1.17450e+02
## 7           1916                3 Non-potable Re-use       2.04680e+02
## 8           1920                1    Potable Treated       6.15700e+03
## 9           1920                2    Non-potable Raw       3.92000e+02
## 10          1920                3 Non-potable Re-use       1.91900e+03
##    metered_water calculated_loss              units
## 1    1.39548e+07     6.57200e+05 Gallons, Thousands
## 2    3.57741e+05     1.02100e+03 Gallons, Thousands
## 3    3.73375e+05    -2.93750e+04 Gallons, Thousands
## 4    6.00114e+08     4.57560e+07            Gallons
## 5    3.97744e+02     3.92170e+01  Gallons, Millions
## 6    1.17445e+02     4.99725e-03  Gallons, Millions
## 7    2.00836e+02     3.84399e+00  Gallons, Millions
## 8    8.85336e+03    -2.69636e+03          Acre-feet
## 9    3.13570e+02     7.84300e+01          Acre-feet
## 10   1.80600e+03     1.13000e+02          Acre-feet
```

1.0 Merging Overview & Normalizing_Pop on ce_annual_ndx & Foundational_09 balance data (to get water provider town name)

- ce_annual_ndx represents the unique water provider that is unique for each provider per year of reporting (which means Denver will have as many ce_annual_ndx numbers as years reported)

- Overview & Normalizing are merged first on ce_annual_ndx, then merged with foundational_09_balance_data on the same variable
- Percentage of Water loss is calculated from the resulting file as calculated_lossS / distributed_watS
- There are about 46 towns where metered water is NA which results in a non-useable percLoss
- Lafayette for 2015, 2016 had no distributed water loss

```
### ce_index is unique for water provider
### ce_annual_ndx is unique for year and provider
joined <- merge(overview, normalizing_pop, by="ce_annual_ndx", all=TRUE)

### remove columns
###
remove <- c("bndss_basin_index", "bndss_type_index", "provider_comments", "report_status", "bndss_id", "pop_served_source",
"transient_pop_source", "remark")
joined <- joined[, !(names(joined) %in% remove)]
head(joined, n=10)
```

```
##    ce_annual_ndx ce_index                                water_provider
## 1           1912      863                                Aurora, City of
## 2           1913      499                                Cortez, City of
## 3           1916      734                              Superior, Town of
## 4           1920      463                   Broomfield, City and County of
## 5           1921      434                                Arvada, City of
## 6           1922      894                          Loveland Water and Power
## 7           1924      714 South Adams County Water and Sanitation District
## 8           1925      692                  Pueblo, Board of Water Works of
## 9           1926      647       Mount Werner Water and Sanitation District
## 10          1927      478               Central Weld County Water District
##           basin     type                          county report_year
## 1         Metro  Utility       Adams, Arapahoe, Douglas        2013
## 2     Southwest  Utility                       Montezuma        2013
## 3  South Platte  Utility                         Boulder        2013
## 4      Colorado  Utility Boulder, Broomfield, Jefferson        2013
## 5         Metro  Utility               Adams, Jefferson        2013
## 6  South Platte  Utility                         Larimer        2013
## 7         Metro Title 32                           Adams        2013
## 8      Arkansas  Utility                          Pueblo        2013
## 9   Yampa/White Title 32                           Routt        2013
## 10 South Platte  Utility                            Weld        2013
##    pop_served transient_pop_served transient_pop_type_desc
## 1      351200                    0        Other (describe)
## 2        8700                   NA      - None Specified -
## 3       12483                   NA      - None Specified -
## 4       60885                30326                  Annual
## 5      109745                    0      - None Specified -
## 6       72846                   NA      - None Specified -
## 7       53958                    0      - None Specified -
## 8      107914                   NA      - None Specified -
## 9        4980                12000                Seasonal
## 10      43000                   NA      - None Specified -
```

```
### Combine balance data/loss with overview (city & year)
balanceM <- merge(joined, foundational_09_balance_data, by="ce_annual_ndx", all=TRUE)
head(balanceM, n=10)
```

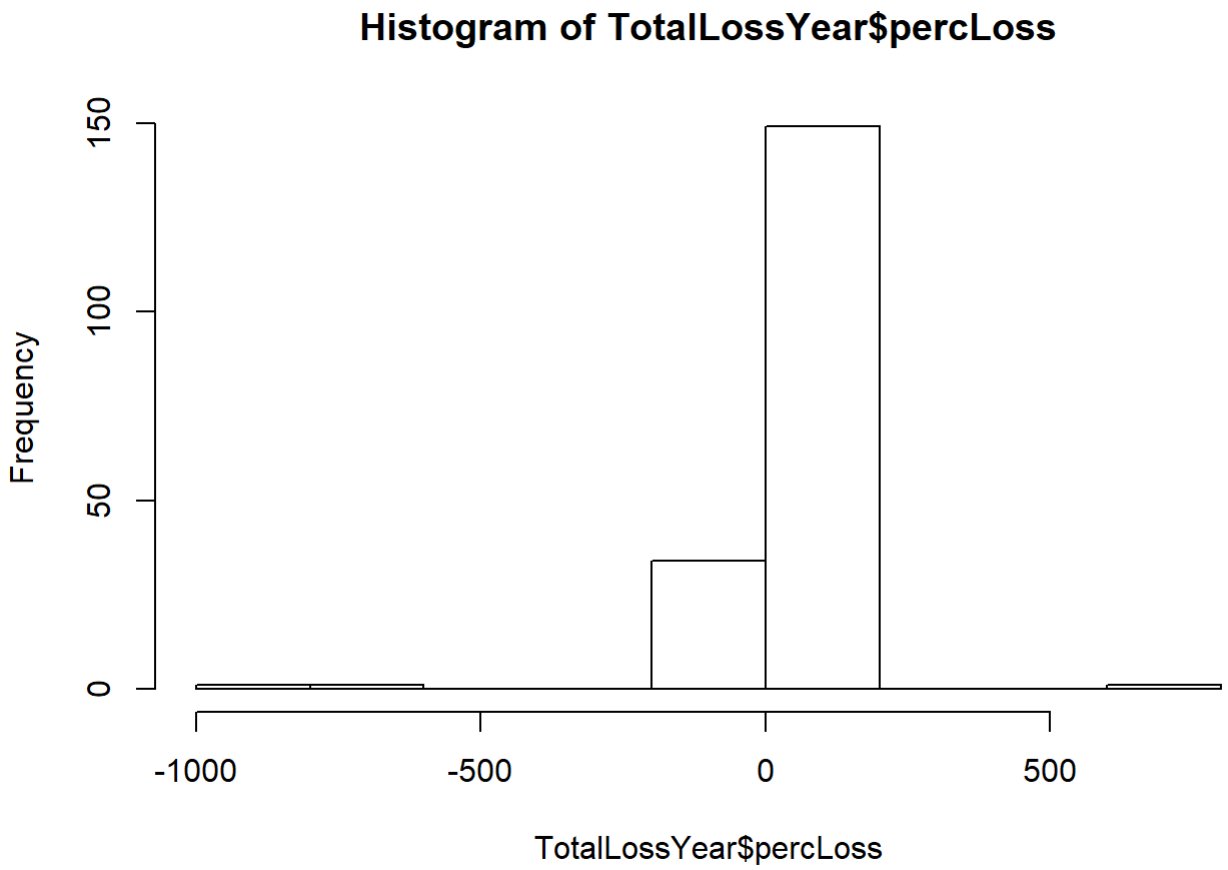
```
##    ce_annual_ndx ce_index                water_provider        basin
## 1           1912      863                Aurora, City of        Metro
## 2           1912      863                Aurora, City of        Metro
## 3           1912      863                Aurora, City of        Metro
## 4           1913      499                Cortez, City of    Southwest
## 5           1916      734              Superior, Town of South Platte
## 6           1916      734              Superior, Town of South Platte
## 7           1916      734              Superior, Town of South Platte
## 8           1920      463 Broomfield, City and County of     Colorado
## 9           1920      463 Broomfield, City and County of     Colorado
## 10          1920      463 Broomfield, City and County of     Colorado
##       type                          county report_year pop_served
## 1  Utility       Adams, Arapahoe, Douglas        2013     351200
## 2  Utility       Adams, Arapahoe, Douglas        2013     351200
## 3  Utility       Adams, Arapahoe, Douglas        2013     351200
## 4  Utility                     Montezuma        2013       8700
## 5  Utility                       Boulder        2013      12483
## 6  Utility                       Boulder        2013      12483
## 7  Utility                       Boulder        2013      12483
## 8  Utility Boulder, Broomfield, Jefferson        2013      60885
## 9  Utility Boulder, Broomfield, Jefferson        2013      60885
## 10 Utility Boulder, Broomfield, Jefferson        2013      60885
##    transient_pop_served transient_pop_type_desc water_type_index
## 1                     0        Other (describe)               1
## 2                     0        Other (describe)               2
## 3                     0        Other (describe)               3
## 4                    NA      - None Specified -               1
## 5                    NA      - None Specified -               1
## 6                    NA      - None Specified -               2
## 7                    NA      - None Specified -               3
## 8                 30326                  Annual               1
## 9                 30326                  Annual               2
## 10                30326                  Annual               3
##            water_type distributed_water metered_water calculated_loss
## 1     Potable Treated       1.46120e+07   1.39548e+07     6.57200e+05
## 2     Non-potable Raw       3.58762e+05   3.57741e+05     1.02100e+03
## 3  Non-potable Re-use       3.44000e+05   3.73375e+05    -2.93750e+04
## 4     Potable Treated       6.45870e+08   6.00114e+08     4.57560e+07
## 5     Potable Treated       4.36961e+02   3.97744e+02     3.92170e+01
## 6     Non-potable Raw       1.17450e+02   1.17445e+02     4.99725e-03
```

```
## 7  Non-potable Re-use       2.04680e+02   2.00836e+02     3.84399e+00
## 8     Potable Treated       6.15700e+03   8.85336e+03    -2.69636e+03
## 9     Non-potable Raw       3.92000e+02   3.13570e+02     7.84300e+01
## 10 Non-potable Re-use       1.91900e+03   1.80600e+03     1.13000e+02
##                 units
## 1  Gallons, Thousands
## 2  Gallons, Thousands
## 3  Gallons, Thousands
## 4             Gallons
## 5   Gallons, Millions
## 6   Gallons, Millions
## 7   Gallons, Millions
## 8           Acre-feet
## 9           Acre-feet
## 10          Acre-feet
```

```
### calcuate sums of water useage by type (potable, non, etc) by city (indx #)
### calculate percent loss = calculated loss/ distributed wats

balanceSum <-balanceM %>% group_by(ce_annual_ndx) %>% summarise(distributed_watS = round(sum(distributed_water, na.rm=TRUE),
 2), metered_waterS=round(sum(metered_water),3),  calculated_lossS= round(sum(calculated_loss), 2) , percLoss = round(calcul
ated_lossS/distributed_watS,2))


### add in descriptor columns
###
c <- c("ce_annual_ndx", "water_provider", "basin", "type", "report_year", "pop_served", "units" )
#colnames(balanceM)

balanceM_short <- balanceM[, (names(balanceM) %in% c)]
together <- merge(balanceSum, balanceM_short, by="ce_annual_ndx")
TotalLossYear <- together[!duplicated(together[,1]),]

# remove all water providers where metered_water is NA
TotalLossYear <- TotalLossYear[!is.na(TotalLossYear$metered_waterS), ]
head(TotalLossYear, n=10)
```

```
##    ce_annual_ndx distributed_watS metered_waterS calculated_lossS percLoss
## 1           1912     1.531476e+07   1.468592e+07        628846.00     0.04
## 4           1913     6.458700e+08   6.001140e+08      45756000.00     0.07
## 5           1916     7.590900e+02   7.160250e+02            43.07     0.06
## 8           1920     8.468000e+03   1.097293e+04         -2504.93    -0.30
## 11          1921     1.445400e+04   1.445420e+04            -0.20     0.00
## 12          1922     4.237200e+09   3.451360e+09     785840000.00     0.19
## 13          1924     2.762583e+06   2.437863e+06        324720.00     0.12
## 15          1925     1.200261e+04   8.170015e+06      -8158011.00  -679.69
## 17          1926     4.720670e+05   4.349490e+05         37118.00     0.08
## 18          1927     9.816100e+03   9.816100e+03             0.00     0.00
##                                    water_provider        basin     type
## 1                                  Aurora, City of        Metro  Utility
## 4                                  Cortez, City of    Southwest  Utility
## 5                                Superior, Town of South Platte  Utility
## 8                    Broomfield, City and County of     Colorado  Utility
## 11                                 Arvada, City of        Metro  Utility
## 12                         Loveland Water and Power South Platte  Utility
## 13 South Adams County Water and Sanitation District        Metro Title 32
## 15                   Pueblo, Board of Water Works of     Arkansas  Utility
## 17       Mount Werner Water and Sanitation District  Yampa/White Title 32
## 18                Central Weld County Water District South Platte  Utility
##    report_year pop_served              units
## 1         2013     351200 Gallons, Thousands
## 4         2013       8700            Gallons
## 5         2013      12483  Gallons, Millions
## 8         2013      60885           Acre-feet
## 11        2013     109745           Acre-feet
## 12        2013      72846             Gallons
## 13        2013      53958 Gallons, Thousands
## 15        2013     107914 Gallons, Thousands
## 17        2013       4980 Gallons, Thousands
## 18        2013      43000           Acre-feet
```

```
hist(TotalLossYear$percLoss)
```




```
# remove Lafayette 2015, 2016
TotalLossYear <- TotalLossYear[!TotalLossYear$distributed_watS ==0,]
```

- The following water providers may have errors in reporting because their metered water is much greater than their distributed water. These were removed from the visualizations below
  - Little Thompson - 2014-2017
  - Pueblo - 2013
  - Lafayette - 2013
  - Louisville - 2017

- 31 rows were removed, see excel chart for which water providers and years

Water Providers With Metered Loss Greater Than Distributed Water, 2013-2017				
2013	2014	2015	2016	2017
Aspen, City of	Aspen, City of	Aspen, City of	Aspen, City of	Aspen, City of
Longmont, City of	Longmont, City of	Longmont, City of	Longmont, City of	Longmont, City of
	Loveland Water and Power	Loveland Water and Power	Loveland Water and Power	Loveland Water and Power
	Superior, Town of	Superior, Town of	Superior, Town of	Superior, Town of
			Parker	Parker
		Centennial		Centennial
	Northglenn, City of	Northglenn, City of	Northglenn, City of	
		Pueblo	Pueblo	
Englewood, City of		Englewood, City of		
Durango, City of		Durango, City of		
Louisville, City of	Westminster, City of	South Adams County	Security Water and Sanitation District	Lafayette, City of
Windsor, Town of	Sterling, City of	Saint Charles Mesa		
Erie, Town of	Grand Junction, City of			

```
TotalLossYear <- TotalLossYear[!(TotalLossYear$metered_waterS > TotalLossYear$distributed_watS),]


# recalculate calculated_loss S column & create
TotalLossYear <- TotalLossYear %>% mutate(calculated_lossS = distributed_watS - metered_waterS , percLoss = round(calculated
_lossS/distributed_watS,2))
```

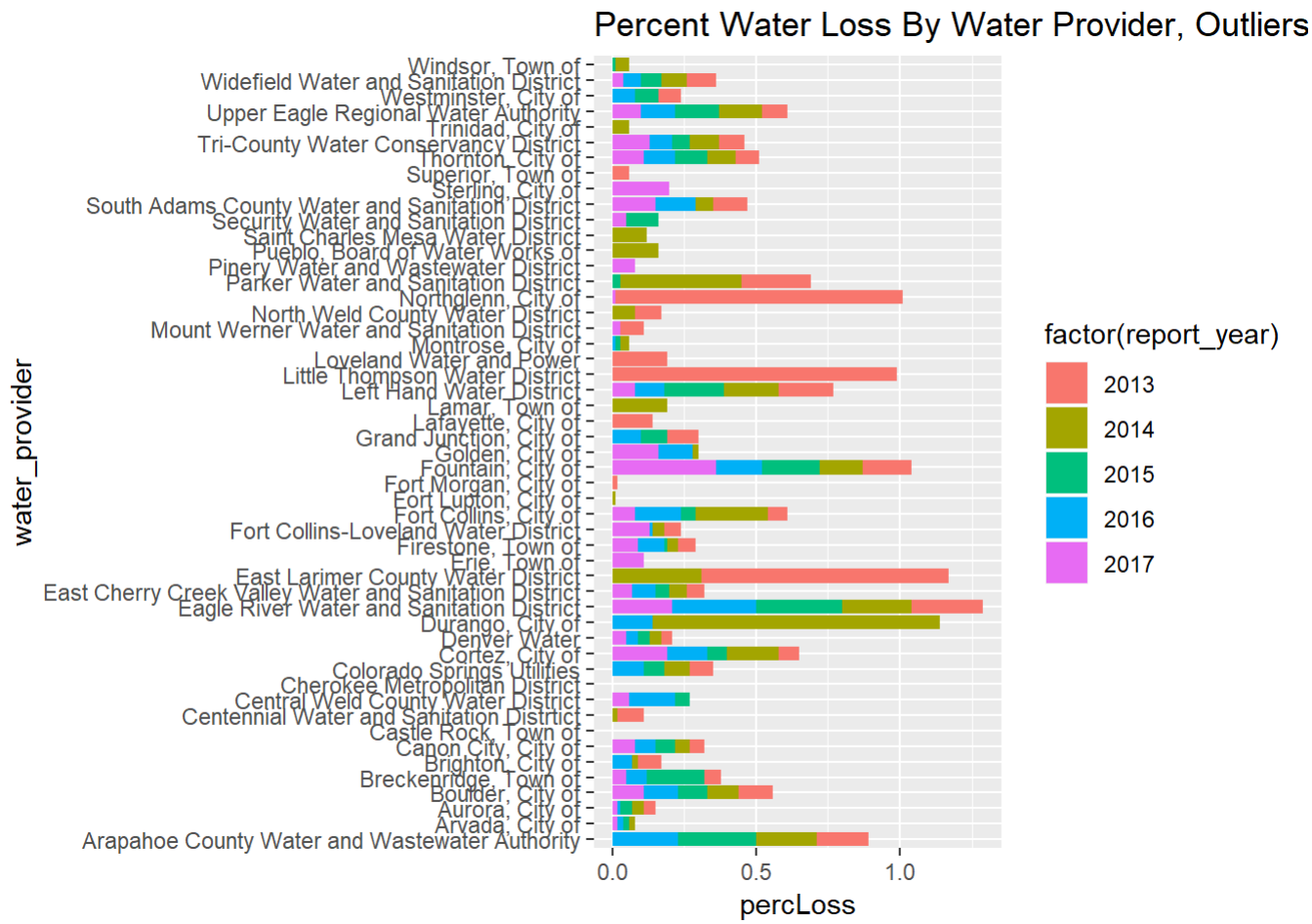
```
TotalLossYear[TotalLossYear$water_provider == "Eagle River Water and Sanitation District", ]
```

```
##     ce_annual_ndx distributed_watS metered_waterS calculated_lossS
## 23           1954           921.16         690.94           230.22
## 43           2001           921.16         696.90           224.26
## 81           2075           971.98         685.20           286.78
## 112          2158           939.00         663.40           275.60
## 150          2223           884.00         697.70           186.30
##     percLoss                            water_provider    basin     type
## 23      0.25 Eagle River Water and Sanitation District Colorado Title 32
## 43      0.24 Eagle River Water and Sanitation District Colorado Title 32
## 81      0.30 Eagle River Water and Sanitation District Colorado Title 32
## 112     0.29 Eagle River Water and Sanitation District Colorado Title 32
## 150     0.21 Eagle River Water and Sanitation District Colorado Title 32
##     report_year pop_served             units
## 23         2013       5289 Gallons, Millions
## 43         2014       5289 Gallons, Millions
## 81         2015       5320 Gallons, Millions
## 112        2016       5450 Gallons, Millions
## 150        2017       5486 Gallons, Millions
```

2.0 Analysis

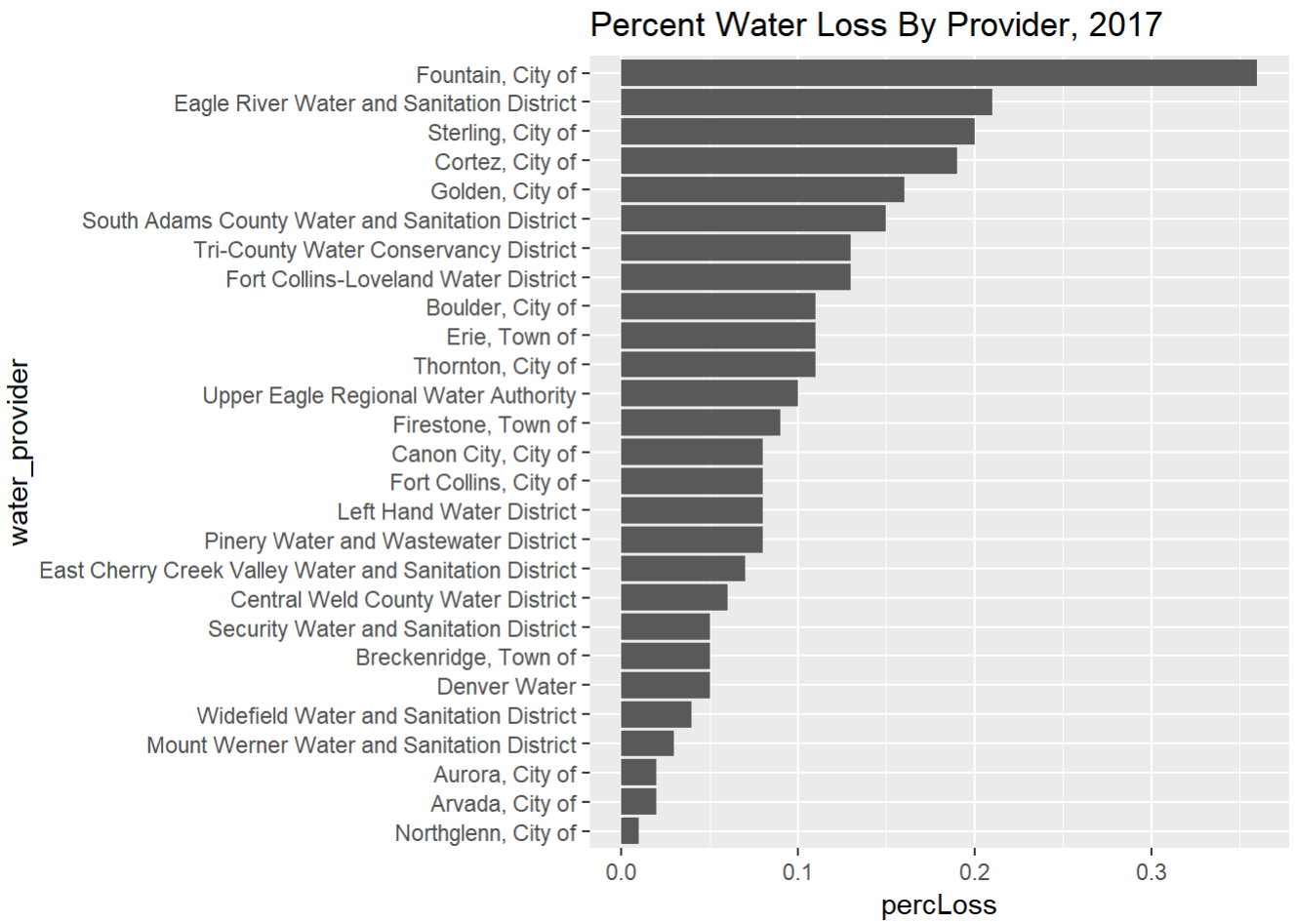
2.1 Plotting Percent Water Loss By Water Provider (Town)

```
####### PLOT
library(scales)

# bar plot by city, by year  - has outliers
ggplot(TotalLossYear, aes(x=water_provider,y=percLoss,  fill=factor(report_year) ))+
  geom_bar(stat="identity") +
  #theme(axis.text = element_text(angle=90)) +
    coord_flip()+
  ggtitle("Percent Water Loss By Water Provider, Outliers Removed, 2013-2017")
```



2.3 ONLY 2017

```
TotalLossYear2017 <- TotalLossYear[TotalLossYear$report_year == 2017,]
#reorder levels
TotalLossYear2017$water_provider <- factor(TotalLossYear2017$water_provider,levels=TotalLossYear2017$water_provider[order(To
talLossYear2017$percLoss)])
ggplot(TotalLossYear2017, aes(x=water_provider,y=percLoss ))+
  geom_bar(stat="identity") +
  # theme(axis.text = element_text(angle=90))+
  coord_flip() +
  ggtitle("Percent Water Loss By Provider, 2017")
```


Percent Water Loss By Provider, 2017
water_provider
Fountain, City of
Eagle River Water and Sanitation District
Sterling, City of
Cortez, City of
Golden, City of
South Adams County Water and Sanitation District
Tri-County Water Conservancy District
Fort Collins-Loveland Water District
Boulder, City of
Erie, Town of
Thornton, City of
Upper Eagle Regional Water Authority
Firestone, Town of
Canon City, City of
Fort Collins, City of
Left Hand Water District
Pinery Water and Wastewater District
East Cherry Creek Valley Water and Sanitation District
Central Weld County Water District
Security Water and Sanitation District
Breckenridge, Town of
Denver Water
Widefield Water and Sanitation District
Mount Werner Water and Sanitation District
Aurora, City of
Arvada, City of
Northglenn, City of
0.0
0.1
0.2
0.3
percLoss