PFAS Levels in North Carolina

https://github.com/kn134/EDAFinal_PFAS.git

Karly Nocera and Tay Holliday

4/26/2021

Contents

List of Tables

List of Figures

```
#sets the default values of chunk options
knitr::opts_chunk$set(echo = FALSE, fig.width = 10, fig.height = 10)
```

- Knitting commands in code chunks:
- `include = FALSE` - code is run, but neither code nor results appear in knitted file
- `echo = FALSE` - code not included in knitted file, but results are
- `eval = FALSE` - code is not run in the knitted file
- `message = FALSE` - messages do not appear in knitted file
- `warning = FALSE` - warnings do not appear...
- `fig.cap = "..."` - adds a caption to graphical results

1 Rationale and Research Questions

The purpose of this study was to synthesize and analyze the findings of local water quality samples measuring a group of emerging contaminants, PFAS (per- and polyfluoroalkyl substances). The two datasets from DEQ (2018 Emerging Compounds Monitoring Reports of various watersheds and public water supply (PWS) reservoirs and the 2019 wastewater treatment plant (WWTP) samples) were chosen as the only publicly-available datasets on local PFAS levels. The sample locations and variables were pre-determined by the study, but these datasets reflect locations and variables pertaining to the levels of various PFAS analytes in the drinking water and water bodies associated with 2.7 million North Carolina residents.

Our study aims to explore and analyze this data in order to begin building the larger story of local contamination and identify areas of concern that need further investigation. To do so, we answered the following questions:

1. Occurence

- What are the most abundant analytes present?
- Which sites have analytes exceeding health advisory levels?
- What chain lengths are most abundant and present?
- What portion of total PFAS are PFOA and PFOS?

2. Distribution

- Are there bodies of water with concerningly high levels of PFAS that aren't commonly discussed?
- How does total PFAS change over time?

3. Unreported / Further Investigation

- What is the highest reporting limit for each analyte?
- How many samples had reporting limits above health advisory levels?
- What percentage of total unreported results were below health advisory limits?

2 Dataset Information

The two datasets were downloaded from North Carolina DEQ: 2018 Analytical Results for PFAS Screening of Select Public Water Supply (PWS) Reservoirs and 2019 publicaly owned utilities with pretreatment programs (POTWs) and industrial dischargers with state permits. Respective to their type of site, both datasets contained site location, sample date, PFAS analyte levels (nanograms per liter (ng/L) or equivalent parts per trillion (ppt)), and lab qualifiers/reporting limits. To substantiate this analysis, we incorporated a manually created csv file identifying the latitude and longitude of sample sites, and researched analytes' chain length (short or long).

Due to the nature of their original format, both datasets needed extensive cleaning prior to importing to R and further exploration and analysis. The POTW data was originally in a PDF and first needed to be converted (using SmallPDF tool) to an excel spreadsheet. These values were then manually moved intoa dataset layout (column headings, rows for each analyte sampled). The PWS dataset was already a csv dataset. For both, analytes with dashes (-) needed to be replaced with a different character (.) in order to read into R and column names were changed to be consistent with other. The lab qualifiers were split so that there was a unique column of numeric reporting limits and in PWS the "results" value was deleted because it was actually the reporting limit and not a result, despite being originally in that column (indicated with corresponding lab qualifier characters).

Once these three datasets were imported into R Studio, both POTW and PWS needed the "sample date" column formatted as a date class and unnecessary columns were removed. To pivot each wider for future use, their reporting limit column was removed and then each pivoted wider so that each analyte became a column with its corresponding parts per trillion measurement. We then mutated both datasets to create columns for Total PFAS, the sum of PFOA and PFOS (due to health advisory regulation), the sum of all short-chained PFAS, and the sum of all long-chained PFAS.

To create a column for chain length in the original long datasets, we used a function to determine if each analyte was short or long, adding the result in the column. Then, both were joined to the imported location dataset so each site had a longitude and latitude. These cleaned files were then saved in the Data»Processed folder as csv.

Data Structure	Value
Variables	Site, Analyte, ppt, Total PFAS, Sum PFOA PFOS, short-chain, long-chain, latitude, longitude
Units	parts per trillion (ppt) or equivalent nanograms per liter (ng/L)
Range	2650 (POTW) and 552 (PWS)
Skew	14.28 (POTW) and 3.56 (PWS)
Kertosis	254.5 (POTW) and 15.2 (PWS)
Links to Data Sources	2018 PWS and 2019 POTWs

3 Exploratory Analysis

3.1 Unique analytes

45 unique analytes were found in POTW samples and 23 in PWS. The following two tables display the most abundant analytes found in each set of site samples.

Table 2: Top 10 Abundant Analytes in POTW

Analyte	Count
PFOA	74
PFHxA	73
PFOS	73
PFHpA	68
PFPeA	54
PFNA	45
PFDA	42
PFHxS	33
PFBS	26
PFBA	25

Table 3: Top 10 Abundant Analytes in PWS

Analyte	Count
PFPeA	17
PFHxA	15
PFHpA	14
PFBA	10
PFOS	8
PFOA	4
PFHxS	2
4.2	0
6.2	0
8.2	0

3.2 Number of samples with each analyte

The following histograms visualize the number of samples of each analyte. Same as above, the abundnace indicates which PFAS are most commonly present. This is relevant for regulation and possibly tracking contaminant sources.

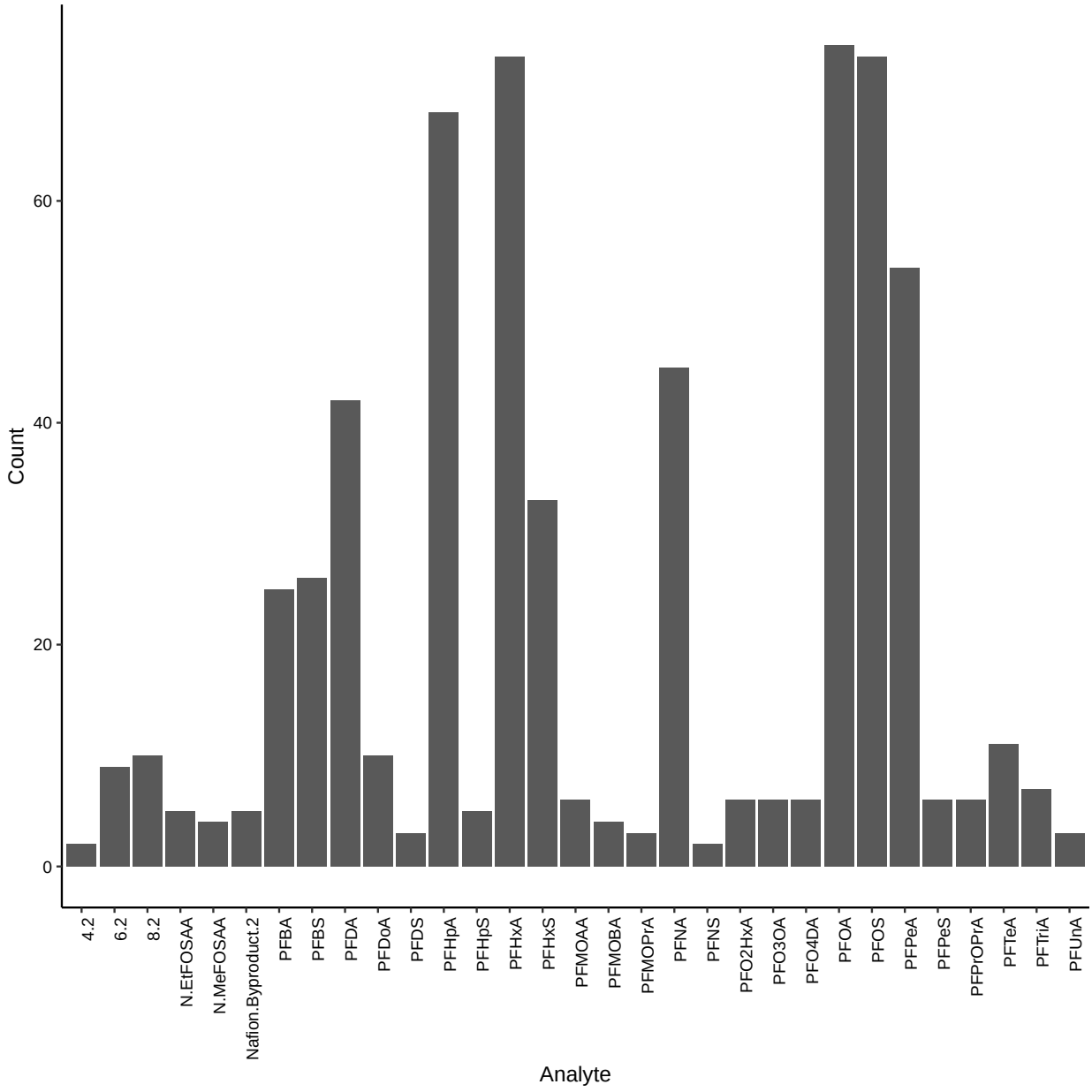


Figure 1: Number of Samples in POTW

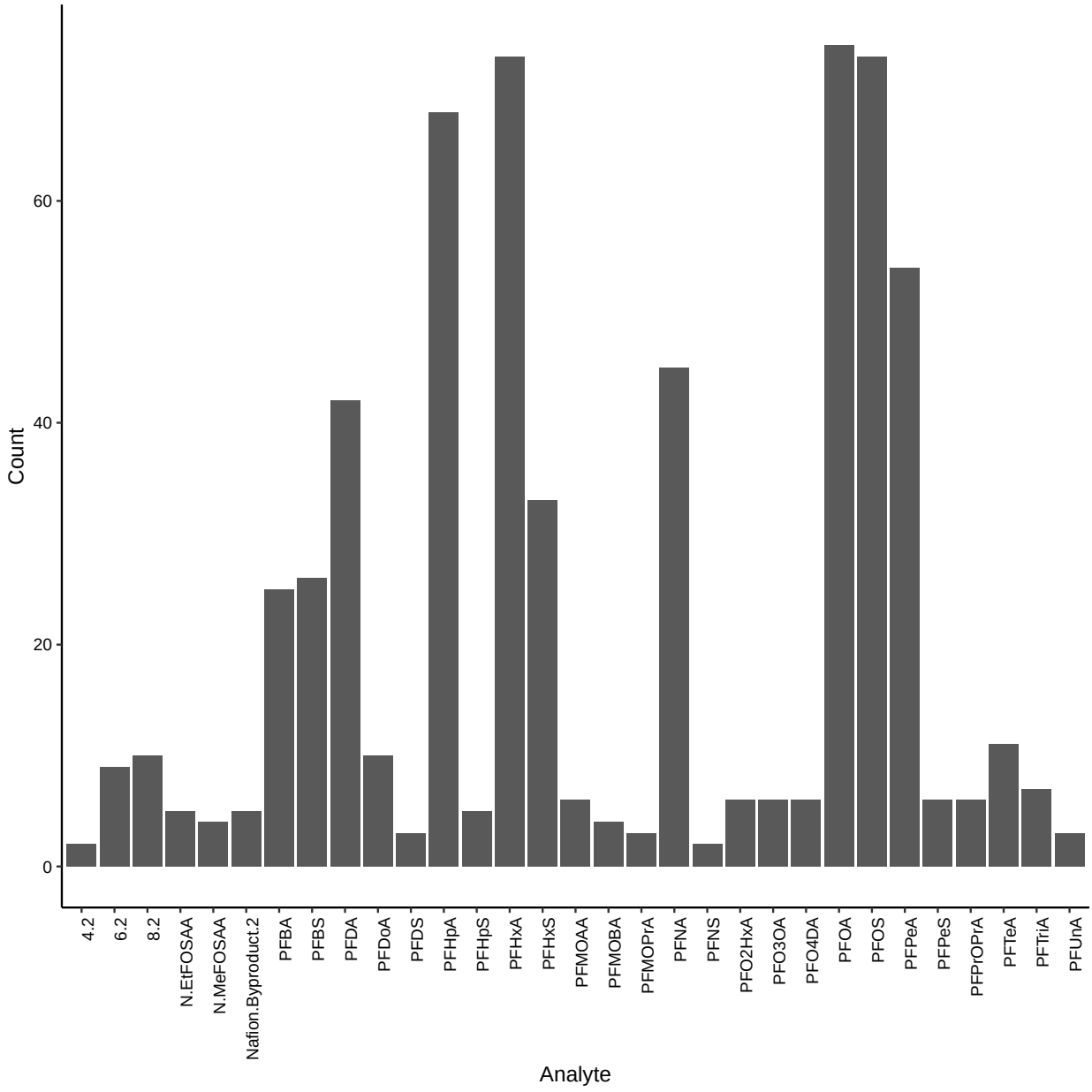


Figure 2: Number of Samples in PWS

3.3 Number of Select Analytes by Level

The following histograms visualize the number of samples of PFAS levels. The abundnace indicates which PFAS are most commonly present and which analytes have a high count at higher levels (ppt). This is relevant for regulation and possibly tracking contaminant sources.

```
## Warning: Removed 237 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 18 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 1209 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 12 row(s) containing missing values (geom_path).
```

3.4 Density ridge lines of select analytes

A closer look the count of analyte samples by value (ppt) indicates which analytes have frequently high values. This is valuable knowledge to begin analyzing whether high levels are an isolated event or consistently high.

```
## Picking joint bandwidth of 9.19
```

```
## Warning: Removed 226 rows containing non-finite values (stat_density_ridges).
```

```
## Picking joint bandwidth of 17.5
```

```
## Warning: Removed 1209 rows containing non-finite values (stat_density_ridges).
```

3.5 Distribution of PFOA and PFOS samples

Violin plots illustrating the distribution of PFOA and PFOS in both sample site types indicate the quartiles of both analytes. This is important for understanding the PFAS levels where the bulk of samples occured.

```
## Warning: Removed 17 rows containing non-finite values (stat_ydensity).
```

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```

```
## Warning: Removed 268 rows containing non-finite values (stat_ydensity).
```

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```

3.6 PFOA/PFOS exceeding health advisory in POTW

These tables shows the three sites that had the sum of PFOA and PFOS over 70ppt (EPA's health advisory level) during the three-month testing samples.

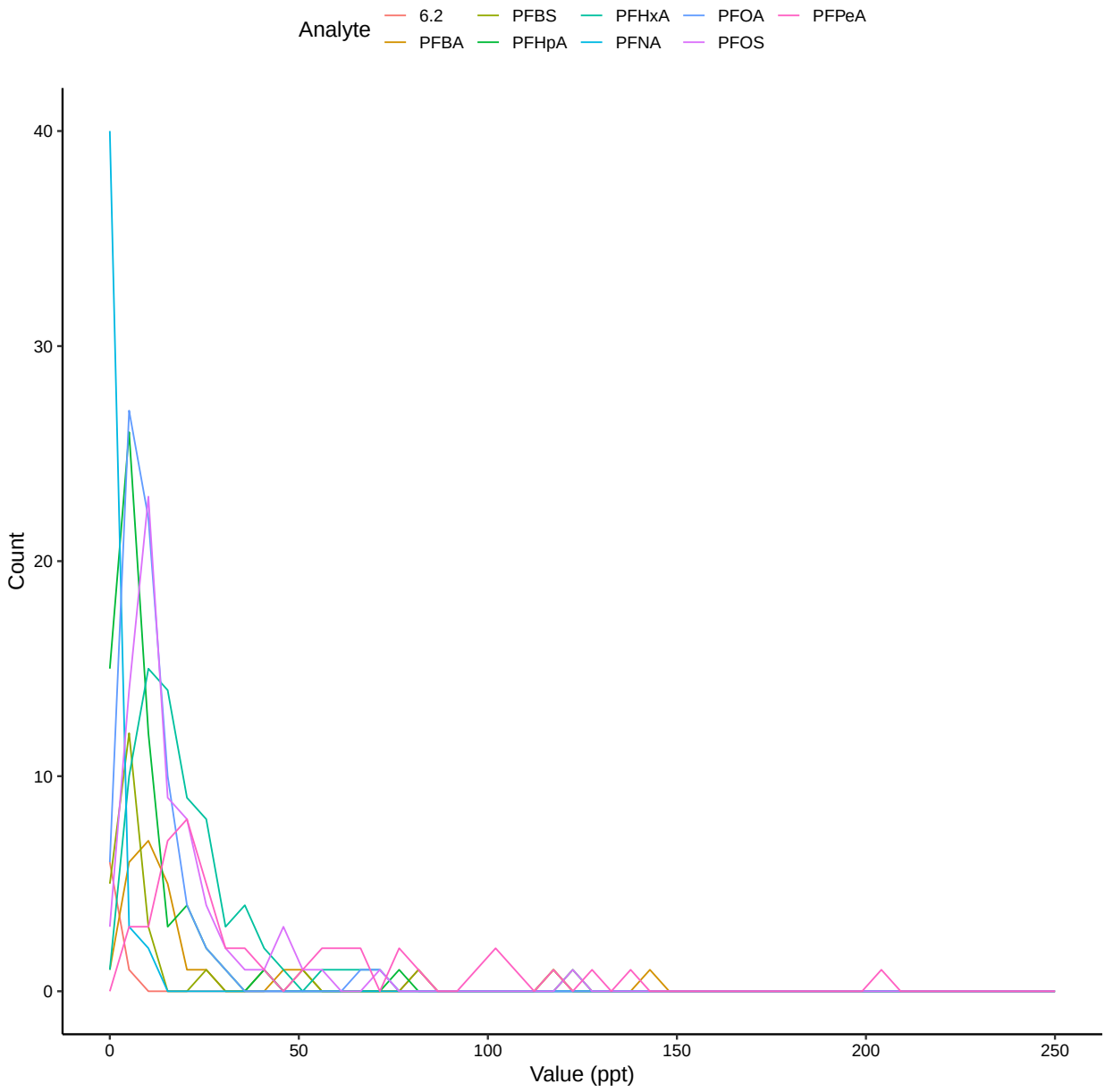


Figure 3: Number of Samples by Level in POTW

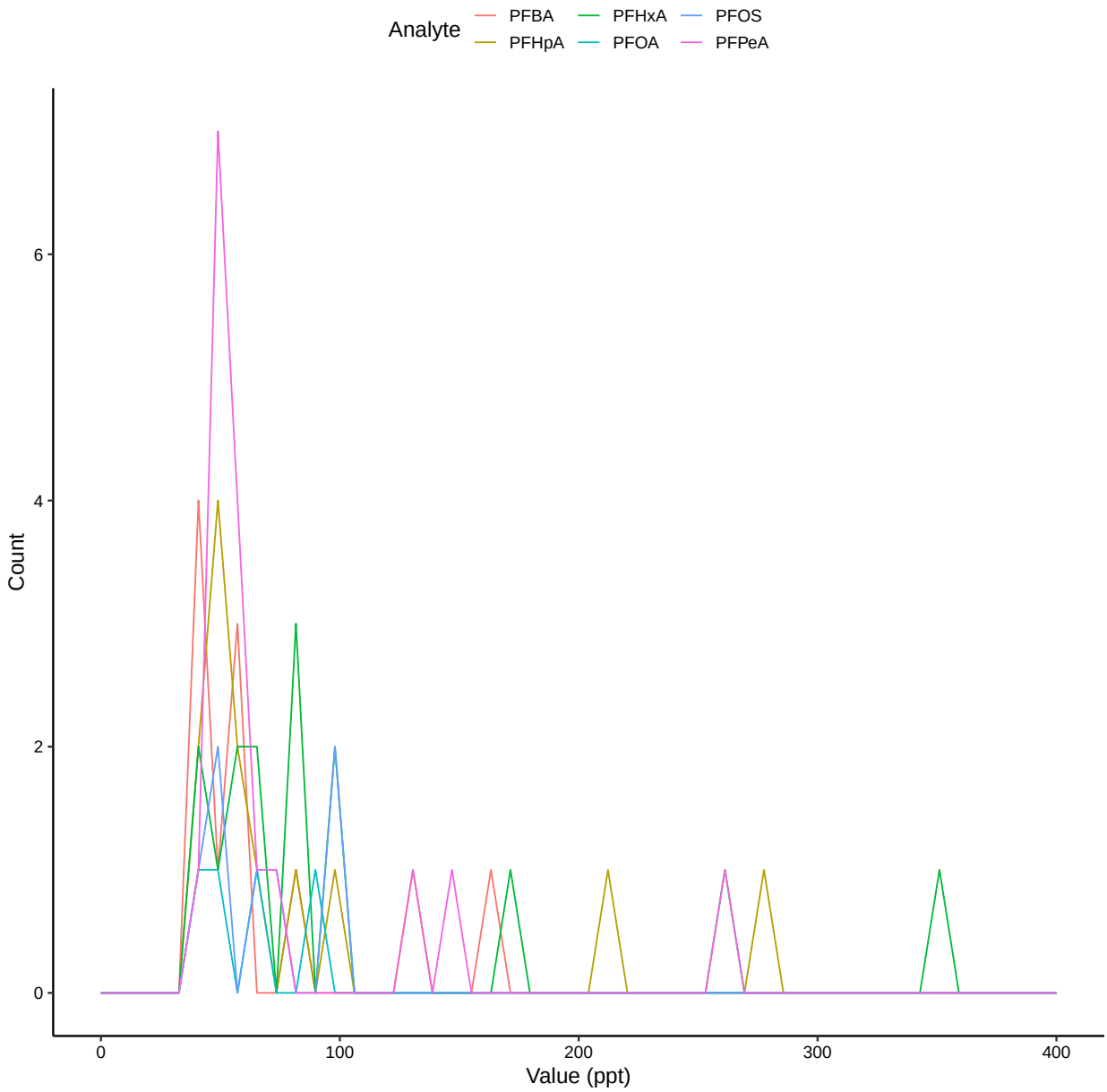


Figure 4: Number of Samples by Level in PWS

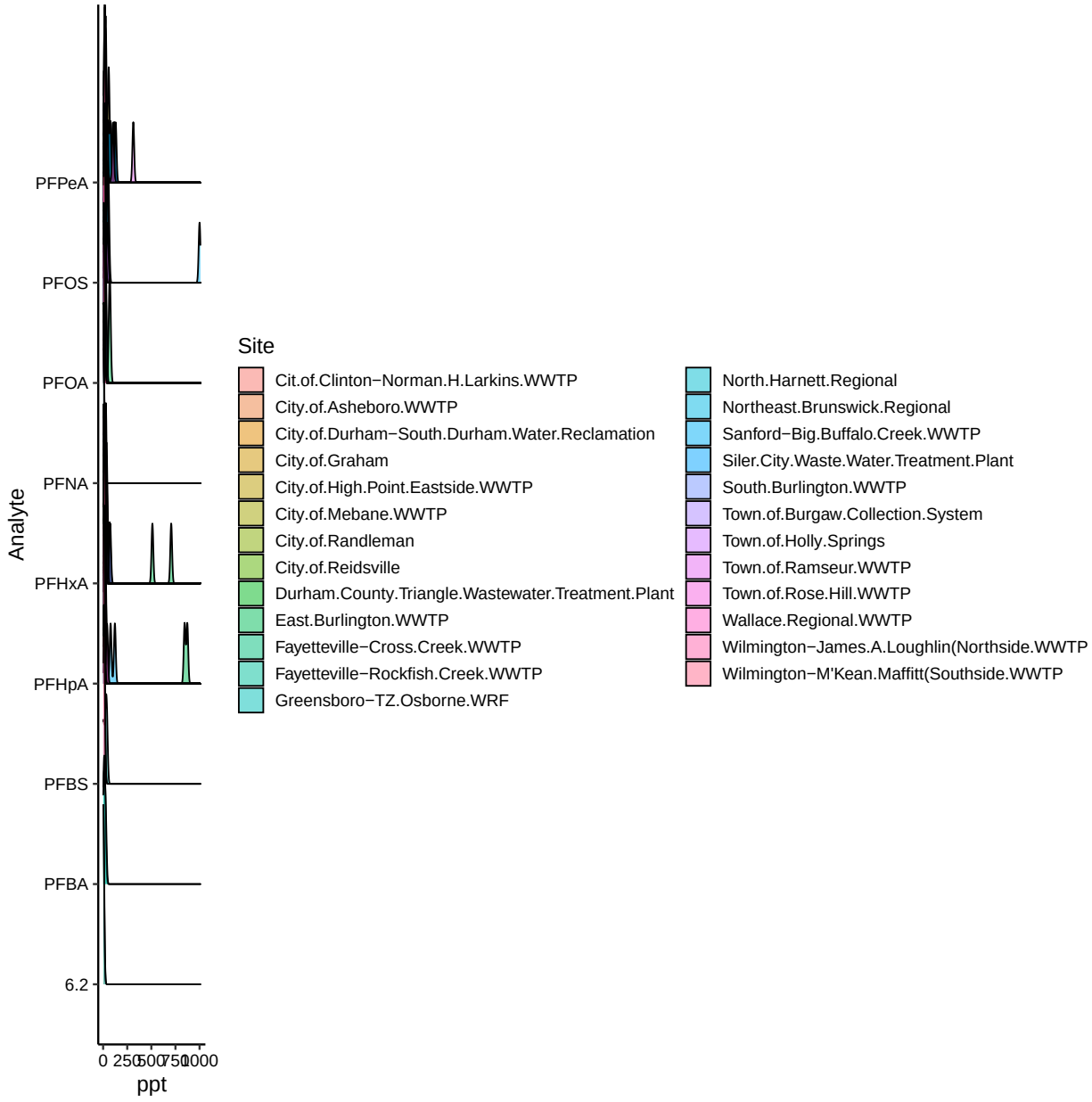


Figure 5: Density Ridges of Select Analytes by Level in POTW

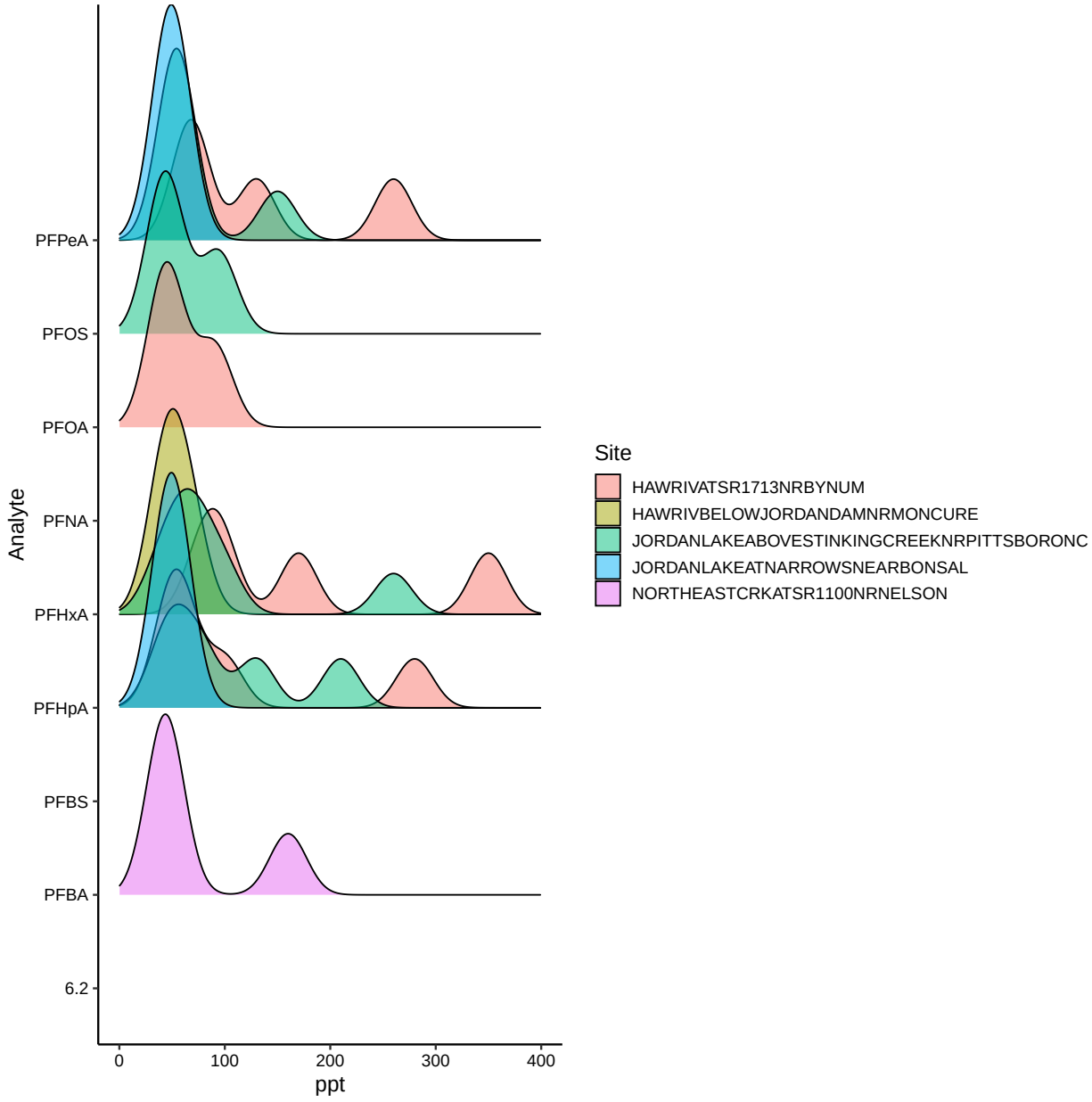


Figure 6: Density Ridges of Select Analytes by Level in PWS

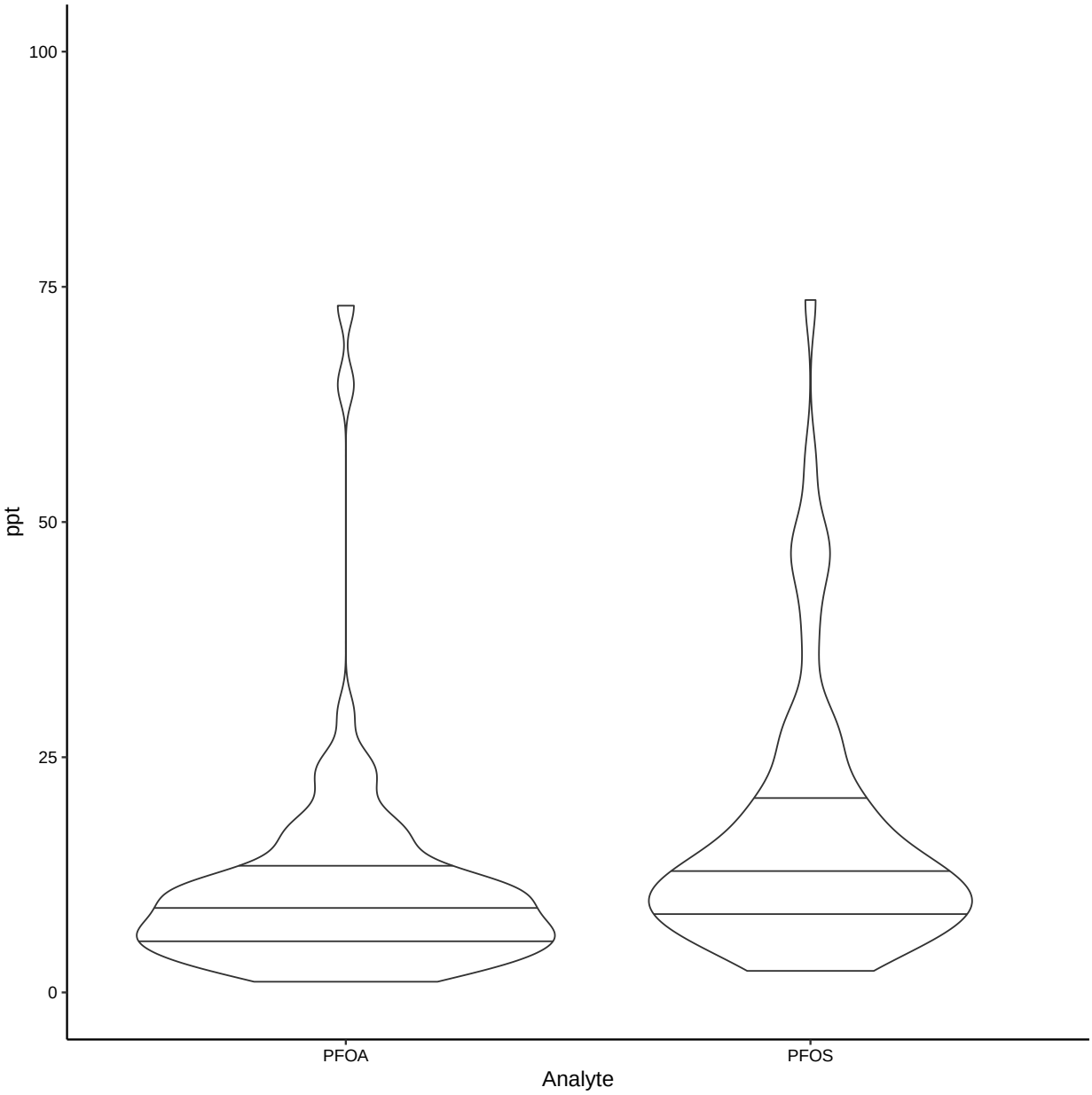


Figure 7: Violin Plots of PFOA and PFOS in POTW

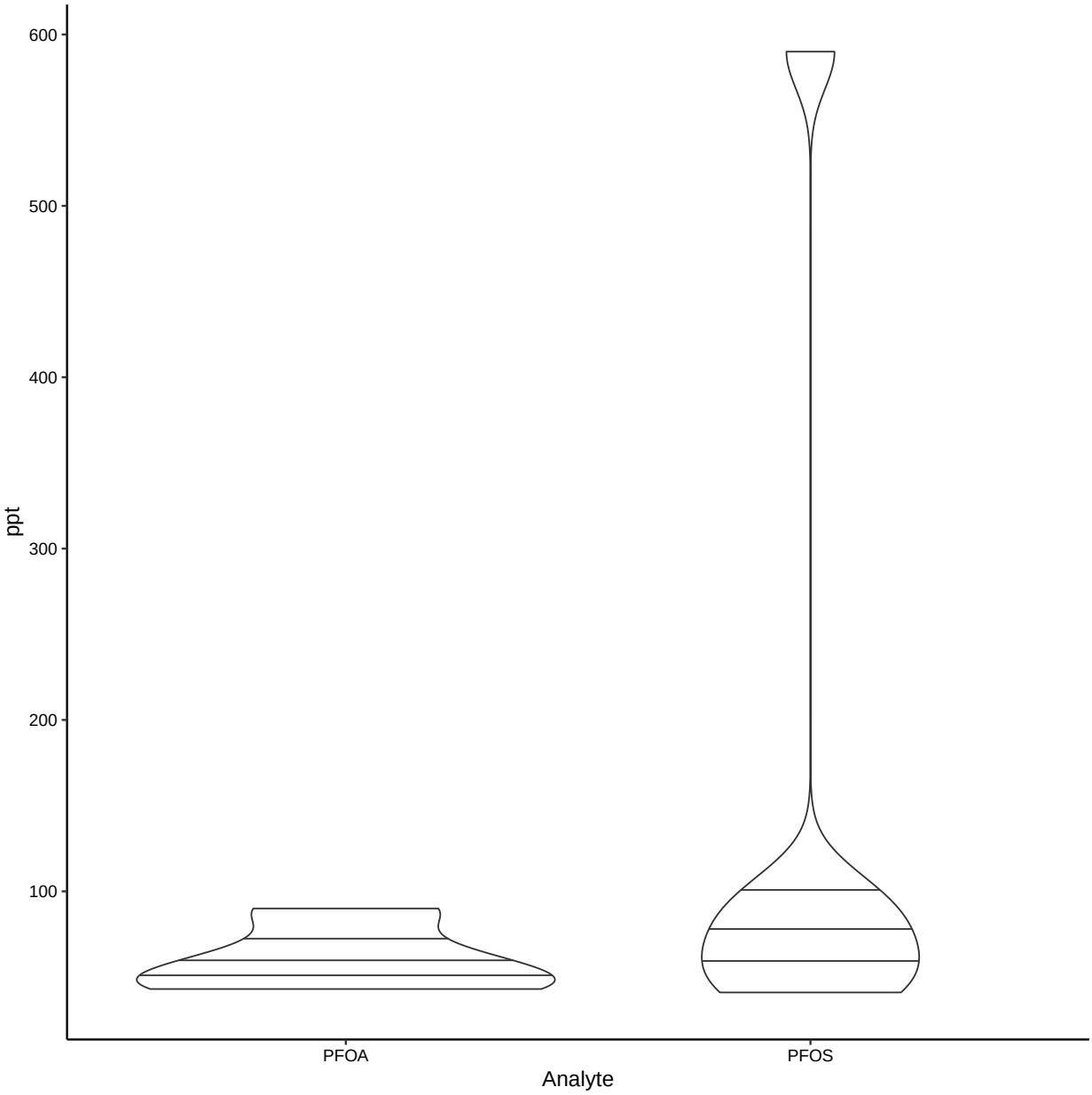


Figure 8: Violin Plots of PFOA and PFOS in PWS

Table 4: Sites Exceeding EPA Health Advisory Limit

Site	PFOA	PFOS	Sum	Sample.Date
City.of.Raeford	NA	124.0	124.0	2019-09-16
City.of.Raeford	NA	73.6	73.6	2019-08-05
City.of.Raeford	NA	NA	0.0	2019-07-15
East.Burlington.WWTP	10.8	12.5	23.3	2019-09-17
East.Burlington.WWTP	64.6	56.4	121.0	2019-08-06
East.Burlington.WWTP	73.0	49.8	122.8	2019-07-16
Sanford-Big.Buffalo.Creek.WWTP	11.0	1000.0	1011.0	2019-09-04
Sanford-Big.Buffalo.Creek.WWTP	11.5	40.8	52.3	2019-08-06
Sanford-Big.Buffalo.Creek.WWTP	11.6	46.3	57.9	2019-07-08

3.7 Maximum reporting limits per analyte

If an analyte was measured below a reporting limit (ppt), it was not reported. Particularly high reporting limits are concerning because it indicates that there are sites that may have high levels of PFAS that are overlooked. While there is no specific value that is "too" high, long-chain PFAS are a known health risk at lower levels (e.g., 8ppt). These tables indicate the maximum reporting limits that occured for each analyte sampled.

Table 5: Max. Reporting Limits of Each Analyte in POTW

Analyte	Max. Reporting Limit
N.EtFOSAA	1900.00
N.MeFOSAA	1900.00
PFDoA	420.00
PFBA	380.00
PFDA	380.00
PFTeA	380.00
PFTriA	380.00
PFUnA	380.00
HFPO.DA	354.62
4.2	333.00
PFHxDA	210.00
PFNA	210.00
PFODA	210.00
PFDS	200.00
PFHpS	170.00
PFNS	170.00
PFOA	170.00
PFOSA	170.00
PFBS	120.00

Analyte	Max. Reporting Limit
PFHpA	120.00
PFHxA	120.00
PFHxS	120.00
PFOS	120.00
PFPeA	120.00
PFPeS	120.00
6.2	89.60
8.2	85.00
PFPrOPrA	21.00
N.MeFOSA	18.90
N.EtFOSA	18.70
N.EtFOSE	9.43
N.MeFOSE	9.43
Nafion.Byproduct.1	3.77
PFECA.G	3.77
PFMOPrA	3.77
10.2	3.62
Nafion.Byproduct.2	3.57
PFMOBA	3.57
PF3OUdS	3.55
PFMOAA	3.40
PFO2HxA	3.40
PFO3OA	3.40
PFO4DA	3.40
ADONA	1.89
PF3ONS	1.76

Table 6: Max. Reporting Limits of Each Analyte in PWS

Analyte	Max. Reporting Limit
6.2	310
PFUnA	310
NMeFOSAA	160
PFBA	160
PFDA	160
PFDoA	160
PFDS	160
PFNA	160
PFNS	160
PFTriA	160
8.2	83
4.2	82

Analyte	Max. Reporting Limit
PFHxA	82
PFOS	80
PFPeA	80
FOSA	42
HFPO.DA	42
PFBS	42
PFHpA	42
PFHpS	42
PFHxS	42
PFOA	42
PFPeS	42

4 Analysis

Each paragraph, accompanied by one or more visualizations, should describe the major findings and how they relate to the question and hypotheses.

4.1 Question 1: Occurence

4.1.1 What are the most abundant analytes present?

FILL IN

```
## Warning: Removed 1303 rows containing missing values (position_stack).
```

```
## Warning: Removed 3166 rows containing missing values (position_stack).
```

4.1.2 Which sites have analytes exceeding health advisory levels?

FILL IN

```
##  [1] City.of.Randleman
##  [2] Wallace.Regional.WWTP
##  [3] South.Burlington.WWTP
##  [4] Fayetteville-Cross.Creek.WWTP
##  [5] Greensboro-TZ.Osborne.WRF
##  [6] City.of.Raeford
##  [7] City.of.High.Point.Eastside.WWTP
##  [8] Town.of.Holly.Springs
##  [9] City.of.Reidsville
## [10] City.of.Mebane.WWTP
## [11] Durham.County.Triangle.Wastewater.Treatment.Plant
## [12] Wilmington-James.A.Loughlin(Northside.WWTP
## [13] Wilmington-M'Kean.Maffitt(Southside.WWTP
## [14] Cit.of.Clinton-Norman.H.Larkins.WWTP
## [15] City.of.Durham-South.Durham.Water.Reclamation
## [16] Sanford-Big.Buffalo.Creek.WWTP
## [17] North.Harnett.Regional
## [18] Town.of.Ramseur.WWTP
## [19] Fayetteville-Rockfish.Creek.WWTP
## [20] Town.of.Rose.Hill.WWTP
## [21] East.Burlington.WWTP
## [22] City.of.Graham
## [23] Northeast.Brunswick.Regional
## 27 Levels: Cary.Western.Wake.Regional.Water.Reclamation.Facility ...
```

```
##  [1] City.of.Randleman
##  [2] Wallace.Regional.WWTP
##  [3] South.Burlington.WWTP
##  [4] City.of.Raeford
```

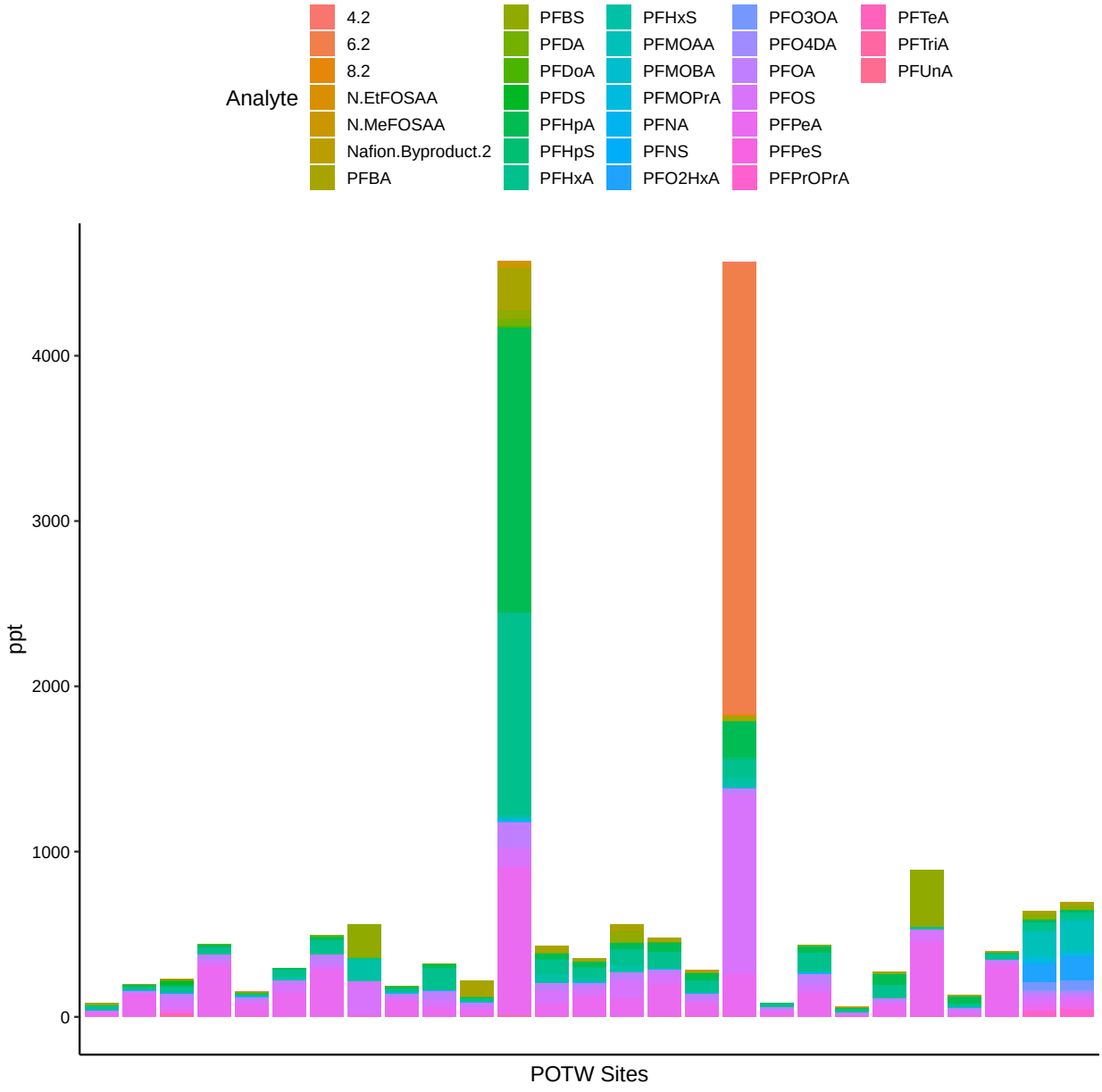


Figure 9: Analytes Present in POTW Sites

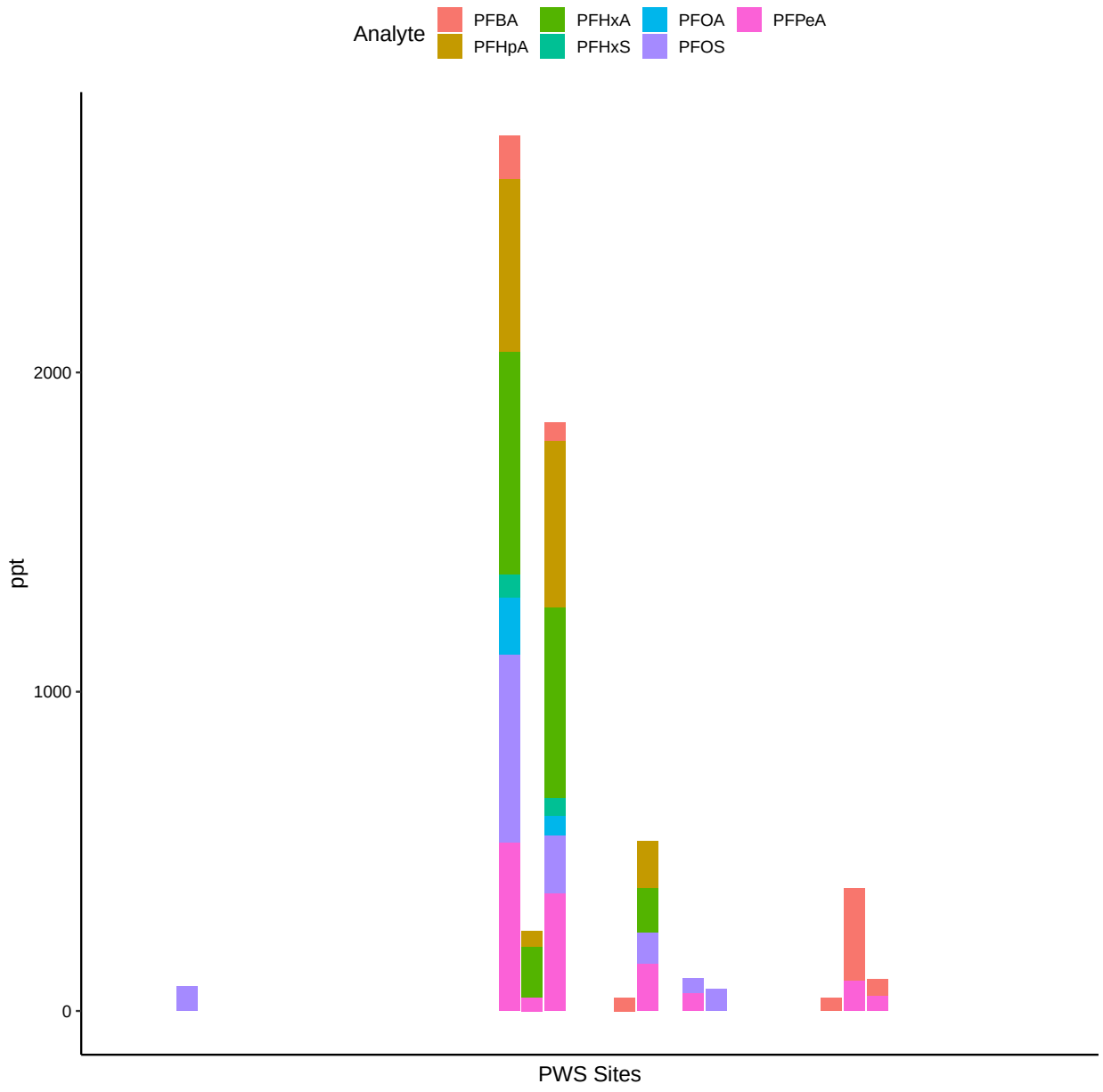


Figure 10: Analytes Present in PWS Sites

```
##  [5] City.of.Mebane.WWTP
##  [6] Cit.of.Clinton-Norman.H.Larkins.WWTP
##  [7] City.of.Durham-South.Durham.Water.Reclamation
##  [8] Sanford-Big.Buffalo.Creek.WWTP
##  [9] North.Harnett.Regional
## [10] Town.of.Ramseur.WWTP
## [11] Fayetteville-Rockfish.Creek.WWTP
## [12] East.Burlington.WWTP
## [13] Northeast.Brunswick.Regional
## 27 Levels: Cary.Western.Wake.Regional.Water.Reclamation.Facility ...
```

```
##  [1] HAWRIVATSR1713NRBYNUM
##  [2] NEWHOPECRKATSR1107NRBLANDS
##  [3] NORTHEASTCRKATSR1100NRNELSON
##  [4] NORTHEASTCRKATSR1731OKELLYCHURCHRDNRDURHAM
##  [5] HAWRIVBELOWJORDANDAMNRMONCURE
##  [6] LakeBrandtatDamnearHillsdale
##  [7] JORDANLAKEABOVESTINKINGCREEKNRPITTSBORONC
##  [8] JORDANLAKEATMOUTHOFWHITEOAKCREEKNRSEAFORTH
##  [9] JORDANLAKEATNARROWSNEARBONSAL
## [10] CaneCreekReservoiratDamnrOaks,NC
## [11] KnapofReedsCreek
## 44 Levels: BlowingRockTownPond BooneASULake ... UniversityLakeatDamnrChapelHill,NC
```

```
## [1] HAWRIVATSR1713NRBYNUM
## [2] NORTHEASTCRKATSR1100NRNELSON
## [3] JORDANLAKEABOVESTINKINGCREEKNRPITTSBORONC
## [4] JORDANLAKEATNARROWSNEARBONSAL
## [5] CaneCreekReservoiratDamnrOaks,NC
## 44 Levels: BlowingRockTownPond BooneASULake ... UniversityLakeatDamnrChapelHill,NC
```

4.1.3 What chain lengths are most abundant and present?

FILL IN

4.1.4 What portion of total PFAS are PFOA and PFOS?

FILL IN

```
## Warning: Removed 210 rows containing missing values (geom_bar).
```

4.2 Question 2: Distribution

4.2.1 Are there bodies of drinking water sources with concerningly high levels of PFAS that aren't commonly discussed?

FILL IN

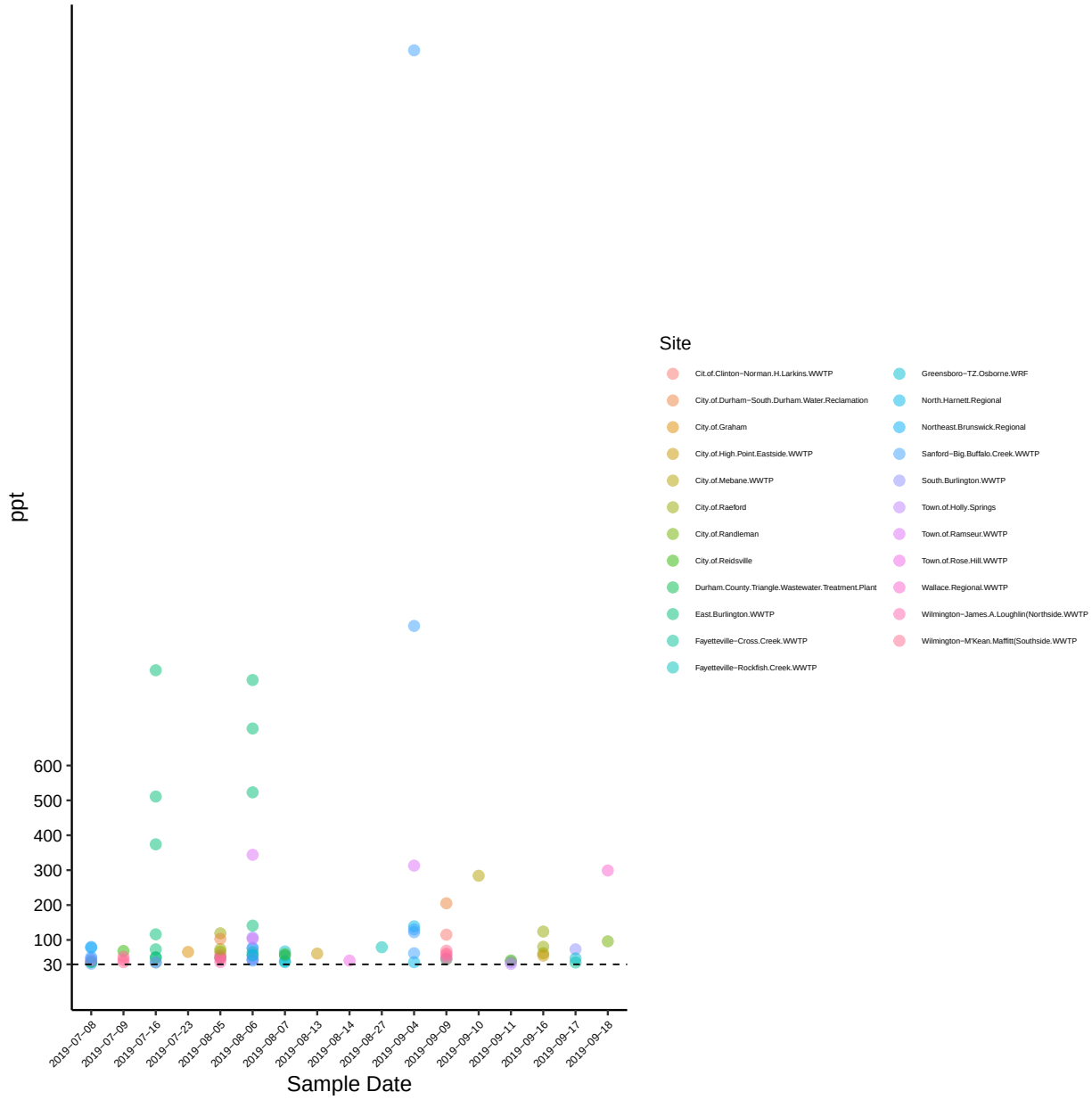


Figure 11: POTW Sites with Analytes Over 30ppt

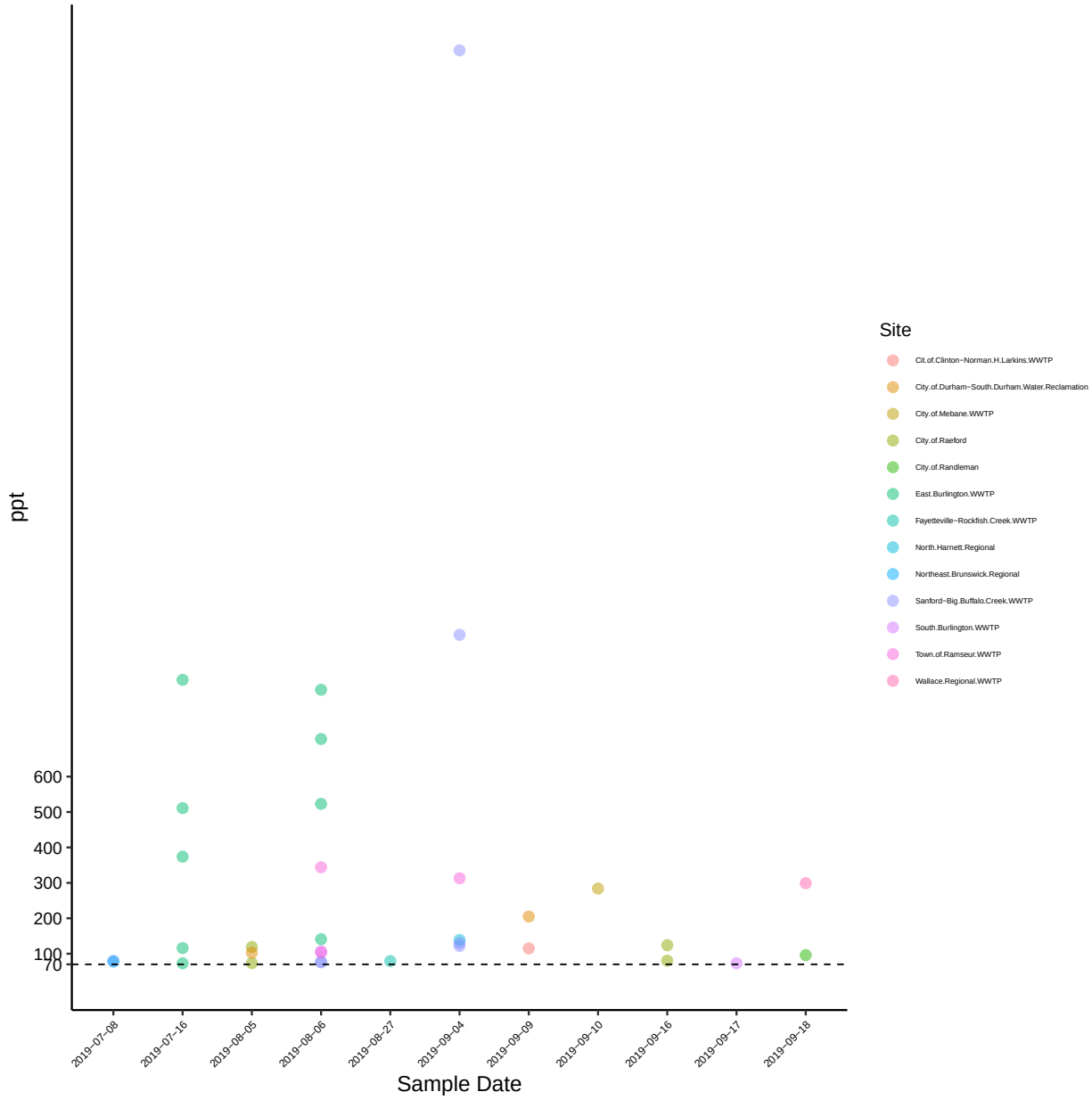


Figure 12: POTW Sites with Analytes Over 70ppt

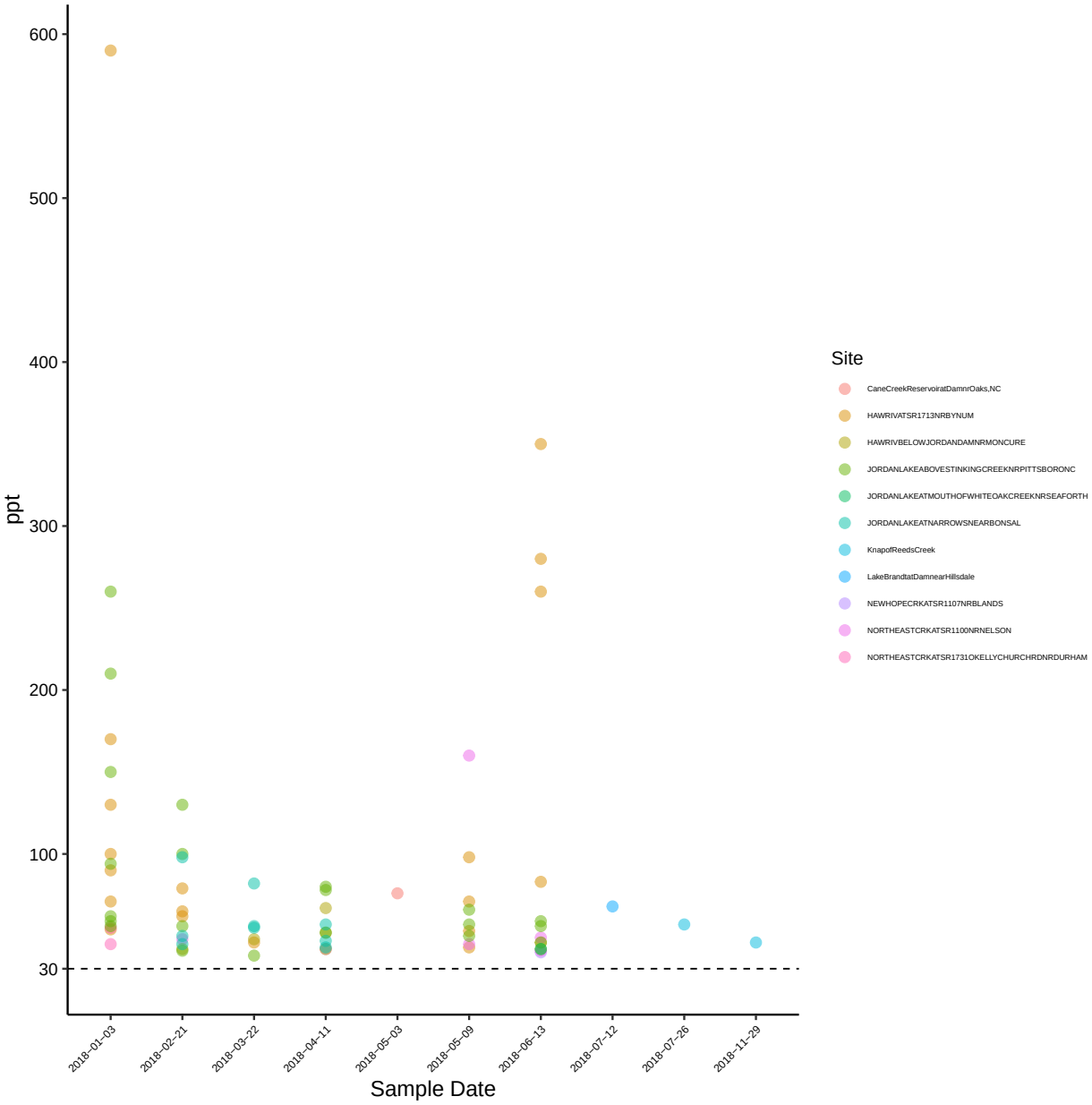


Figure 13: PWS Sites with Analytes Over 30ppt

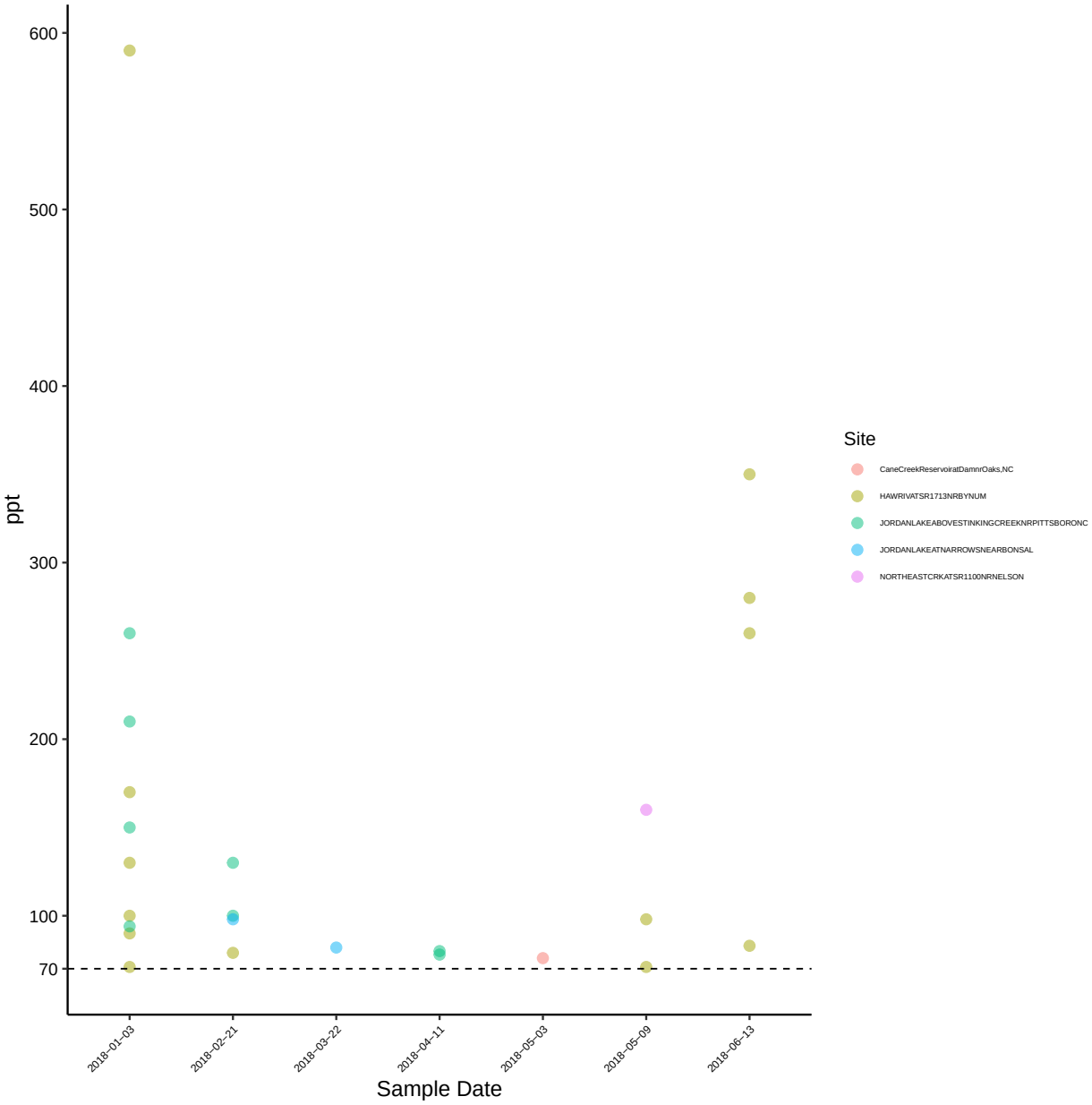


Figure 14: PWS Sites with Analytes Over 70ppt

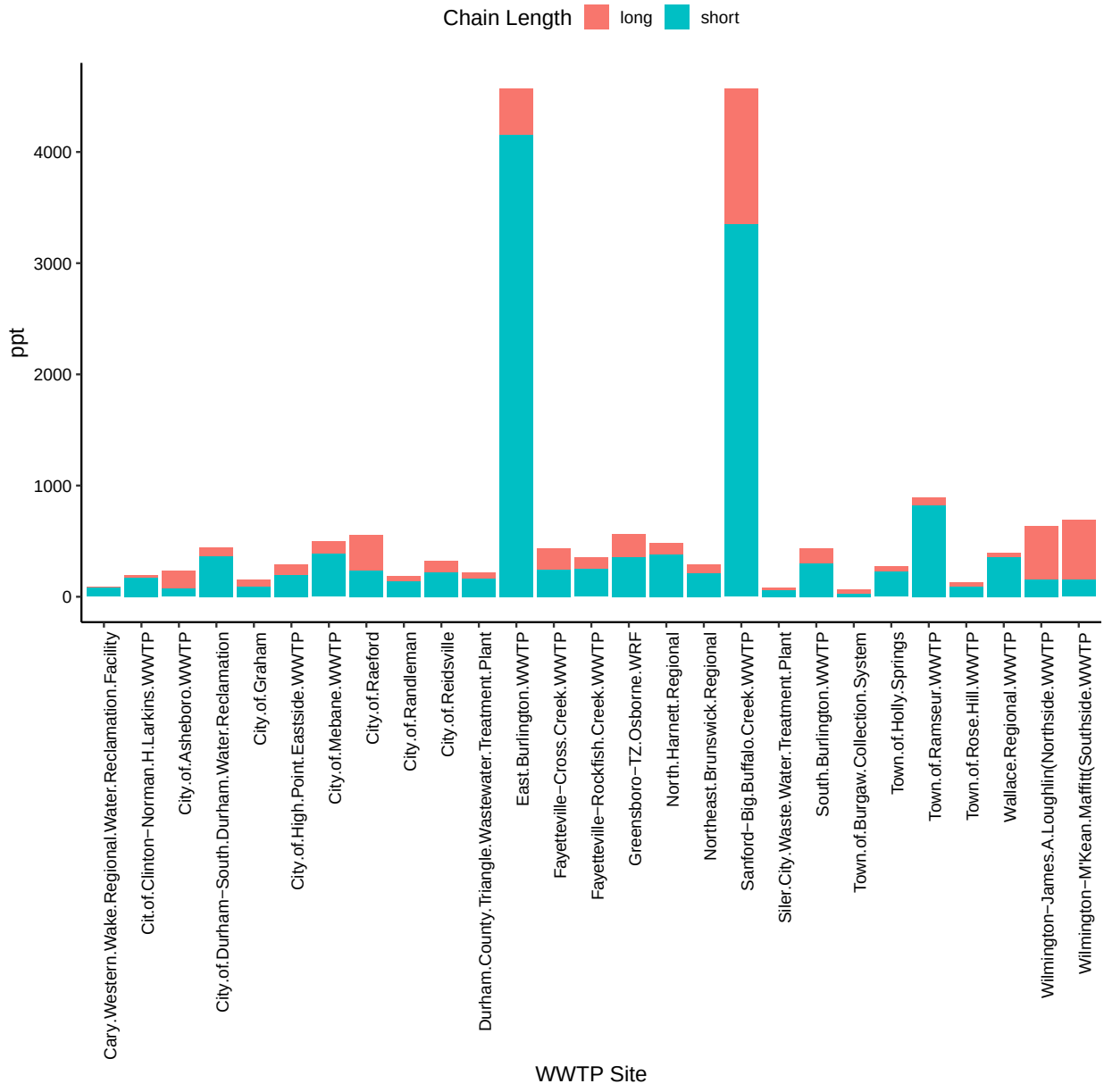


Figure 15: Total PFAS by Chain Length in POTW

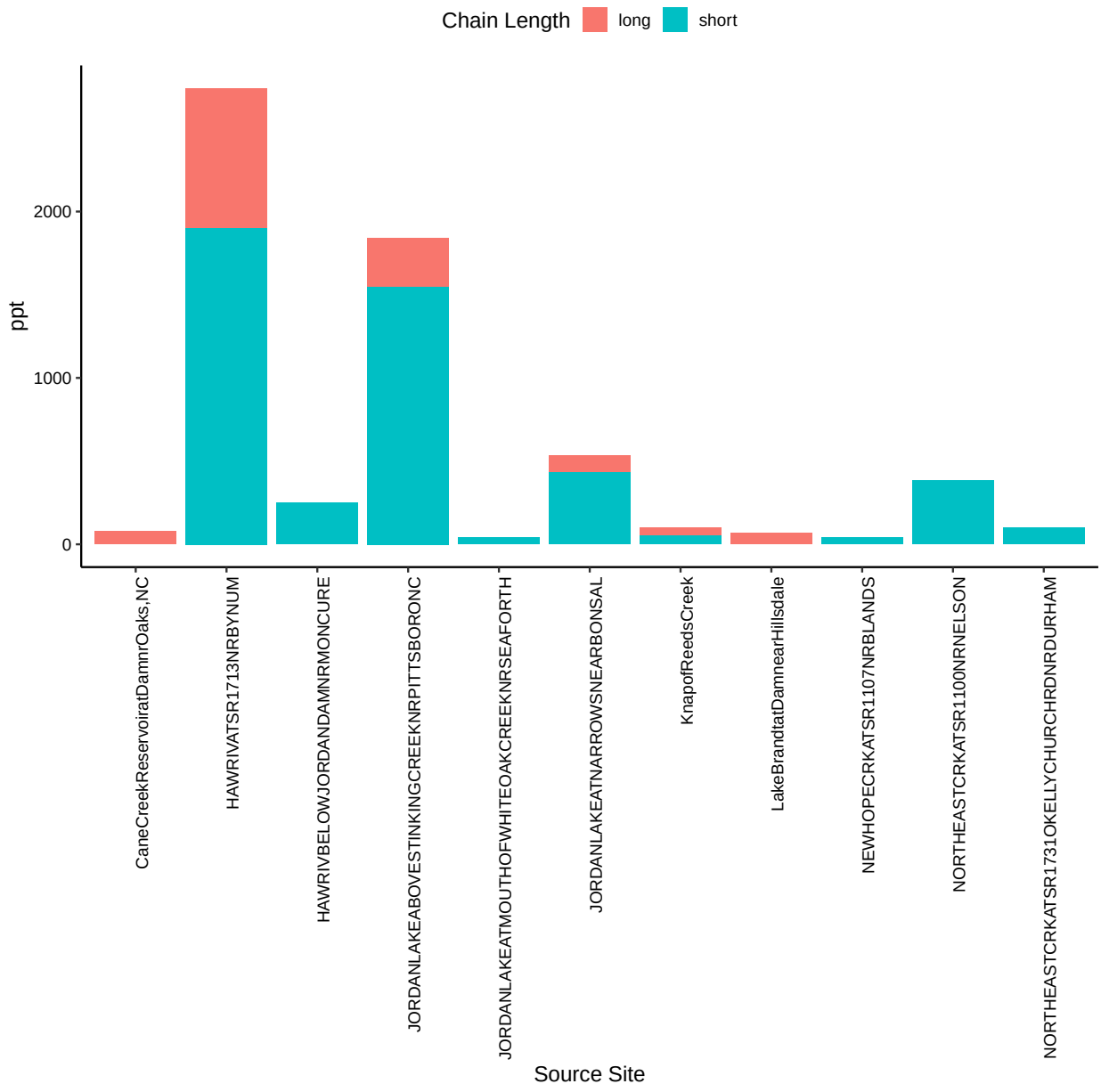


Figure 16: Total PFAS by Chain Length in PWS

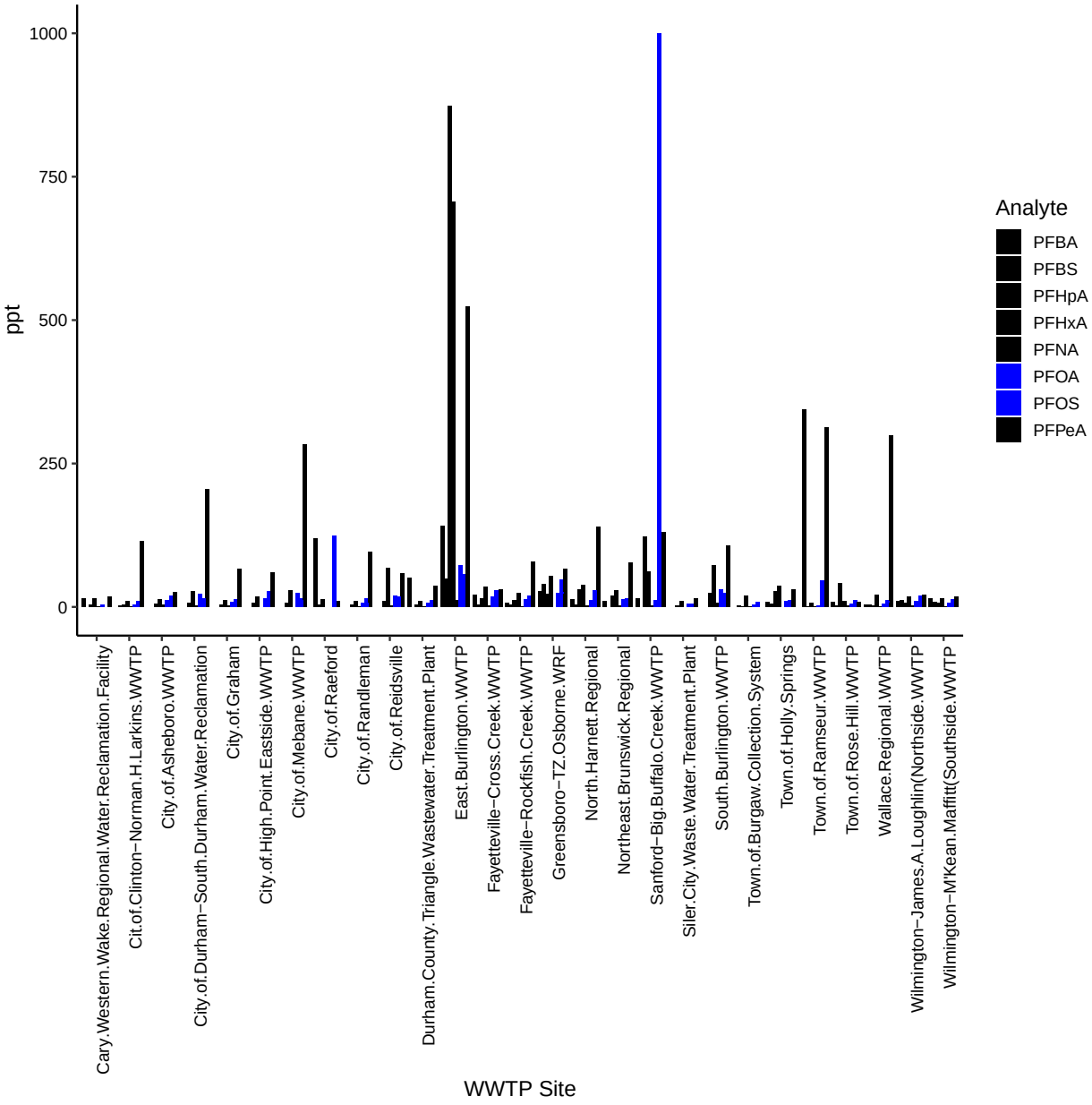


Figure 17: Proportion of PFOA and PFOS in Total PFAS at POTW Sites

NEED TO BE ABLE TO SUM SOURCE DATASET

4.2.2 How does total PFAS change over time?

FILL IN

```
## Warning: Removed 3 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```

4.3 Question 3: Further Investigation

4.3.1 How many samples had reporting limits above health advisory levels?

FILL IN

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

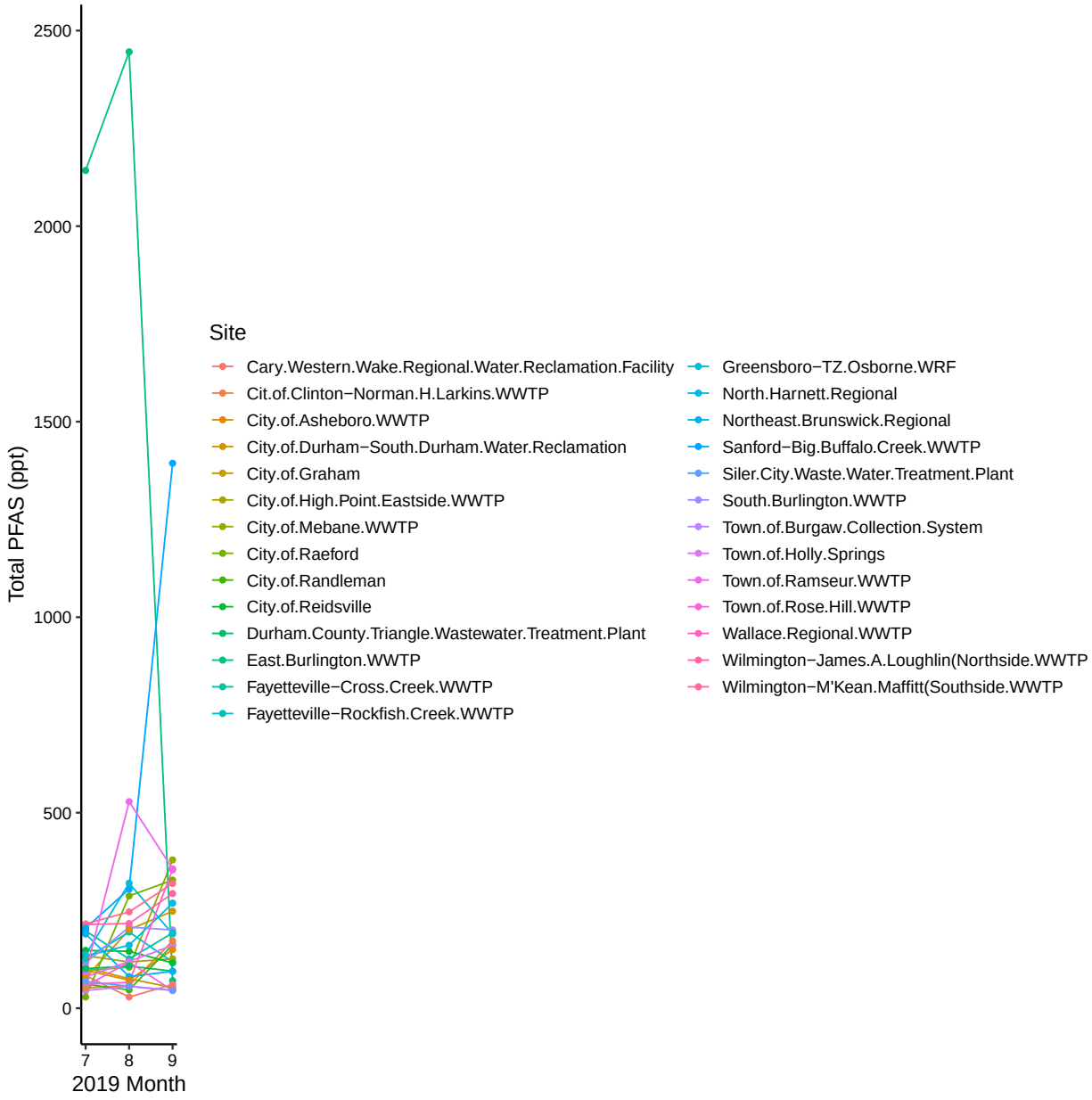


Figure 18: Total PFAS Change Over Time in POTW

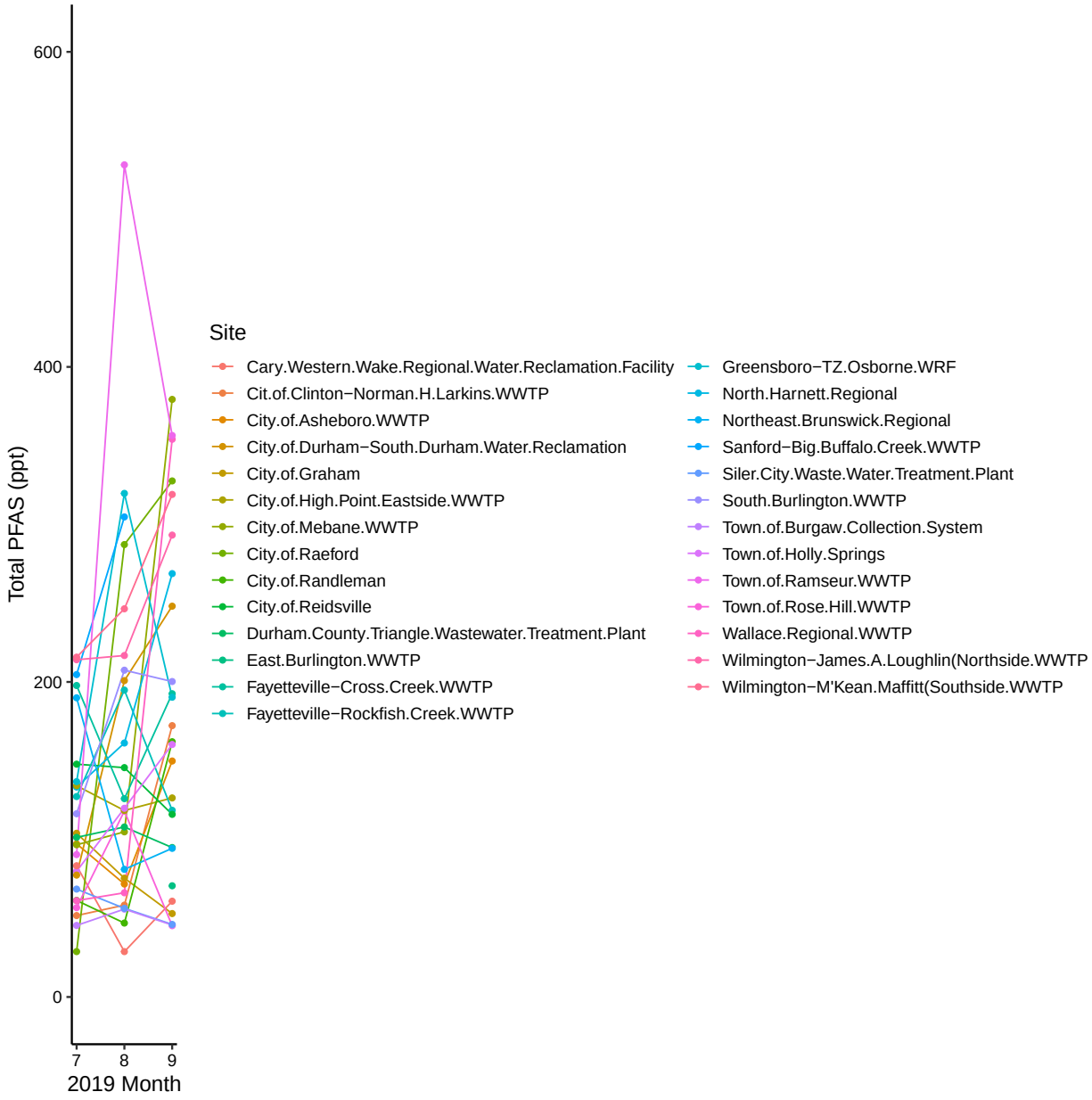
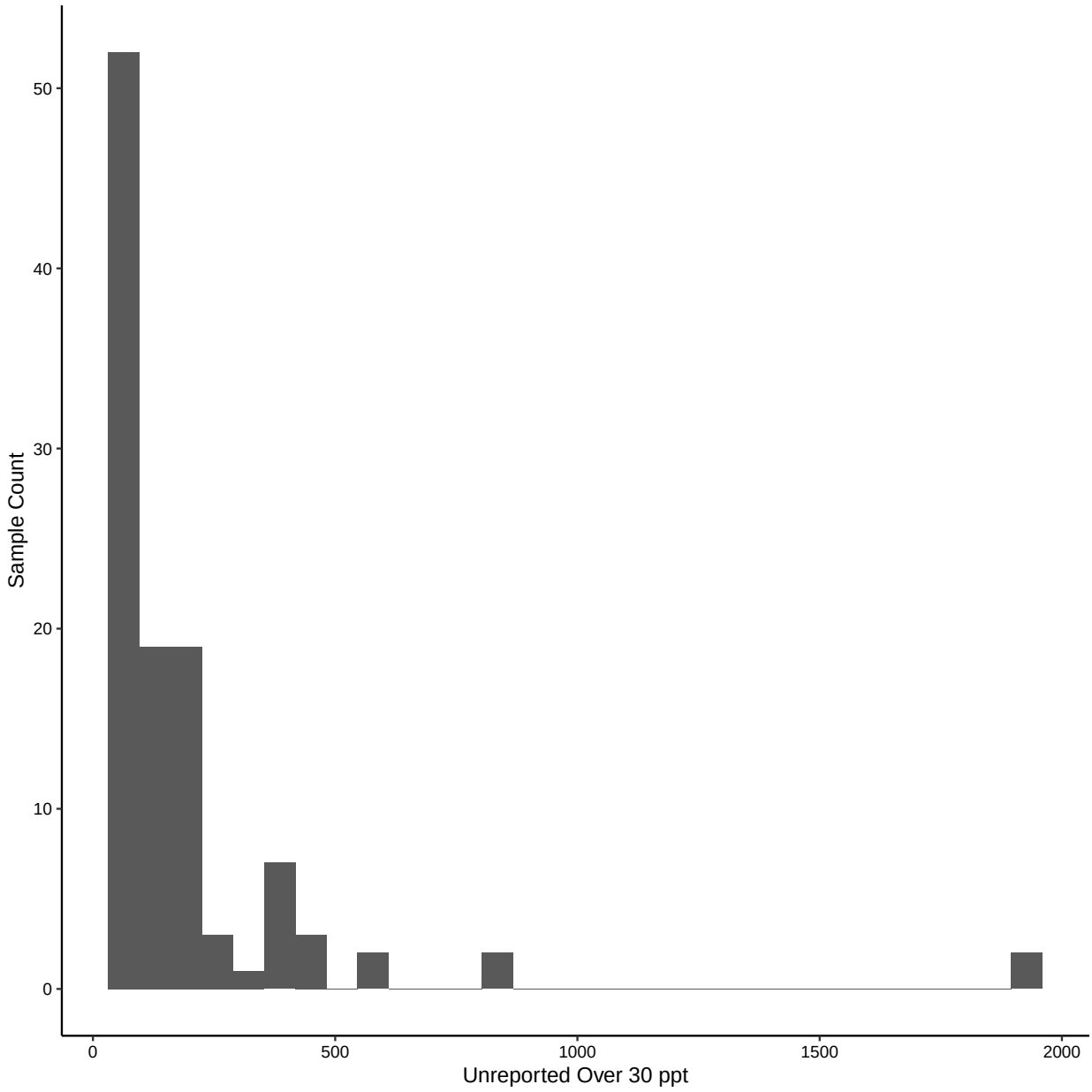


Figure 19: Total PFAS Change Over Time in Select POTW Sites



4.3.2 What percentage of total unreported results were below health advisory limits?

FILL IN

```
## [1] 79.18983
```

```
## [1] 92.29547
```

```
## [1] 0.06317119
```

```
## [1] 84.04927
```

5 Summary and Conclusions

Summarize your major findings from your analyses in a few paragraphs. What conclusions do you draw from your findings? Relate your findings back to the original research questions and rationale.