Using of Cluster Analysis
for
Organizing WRTDS Trend Results.

Elgin Perry, Qian Zhang, Jimmy Webber
USGS Factors Group
4 June 2021

Create a 3-d lattice data structure with dimensions for year, season, and station.

stations included:

1503000 1515000 1531500 1536500 1540500 1576000 1578310 1542500 1549760 1553500 1549700 1502500 1529500 1531000 1534000 1555000 1562000 1567000 1568000 1570000 1571500 1573560 1574000 1576754 1576787 1578475 1580520

months included:

10 11 12 1 2 3 4 5 6 7 8 9
Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep

Years Included

2009 - 2018

Cluster Analysis groups **Items** according to the similarity of the **Profiles.** With the 3-d lattice, it is possible to address a large variety of questions by defining the Items and Profiles differently.

Item	**Profile**	**Question**
Station	Years (average over month)	How does long term trend compare among stations?
Station x Month	Years	Is long term trend at a station consistent over months?
Years	Months (average stations)	Is seasonality changing over time?
Station x Years	Months	Is seasonality at a station consistent over time?
Years	Stations (average months)	Is the longitudinal profile along a river changing over time?

Data Scaling

Scale		**Effect on clusters**
Original	X	Clusters determined by mean level of profile
Mean Adjusted	X – mean(X)	Clusters determined by shape over profile
Z- score	(X – mean(X)) / SD	Clusters determined by relative shape over profile.

1. Cluster analysis grouping items defined by station using patterns over year for S TN observed in the Susquehanna.

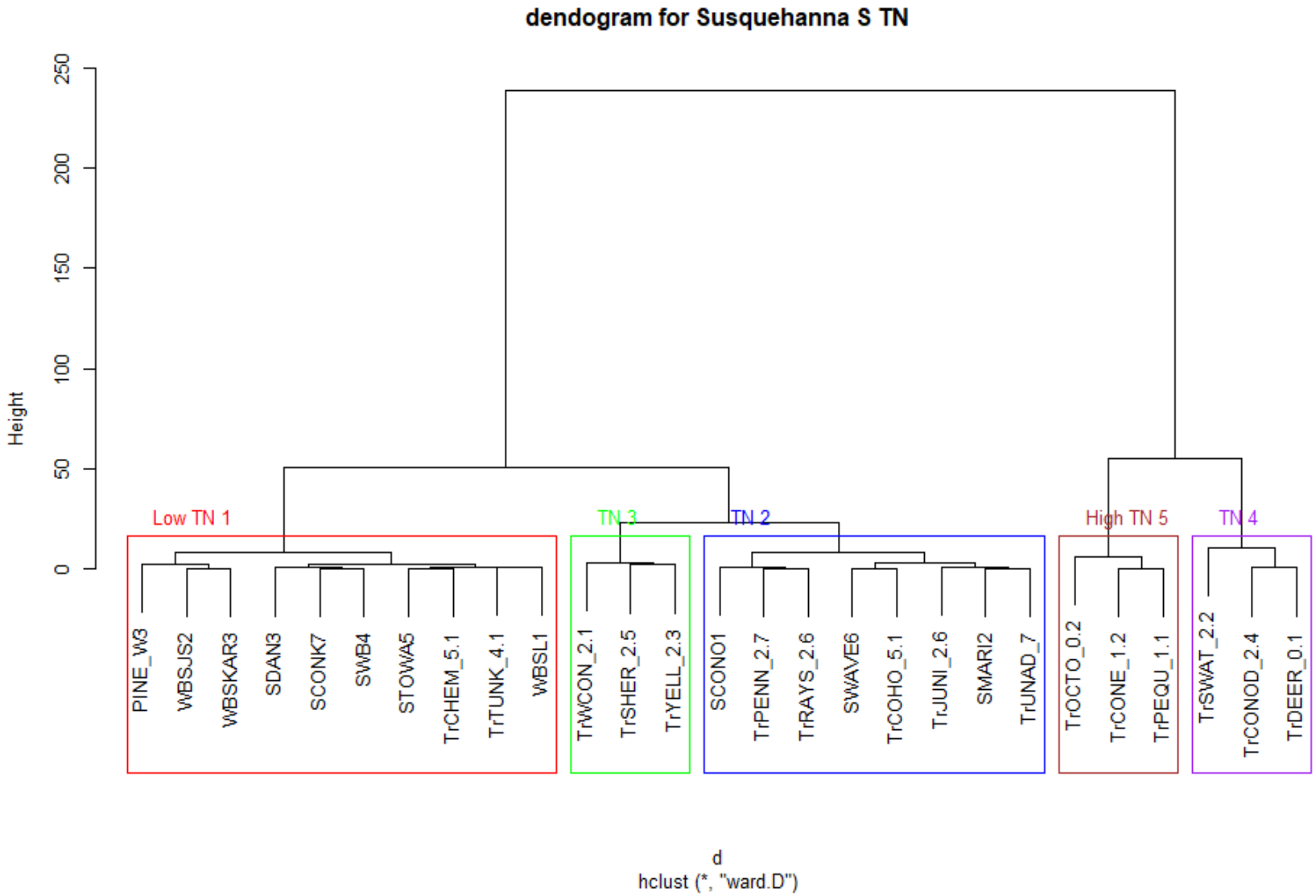


Figure 2.A: Dendogram showing distances between station pairs computed using the dist() function for Susquehanna.

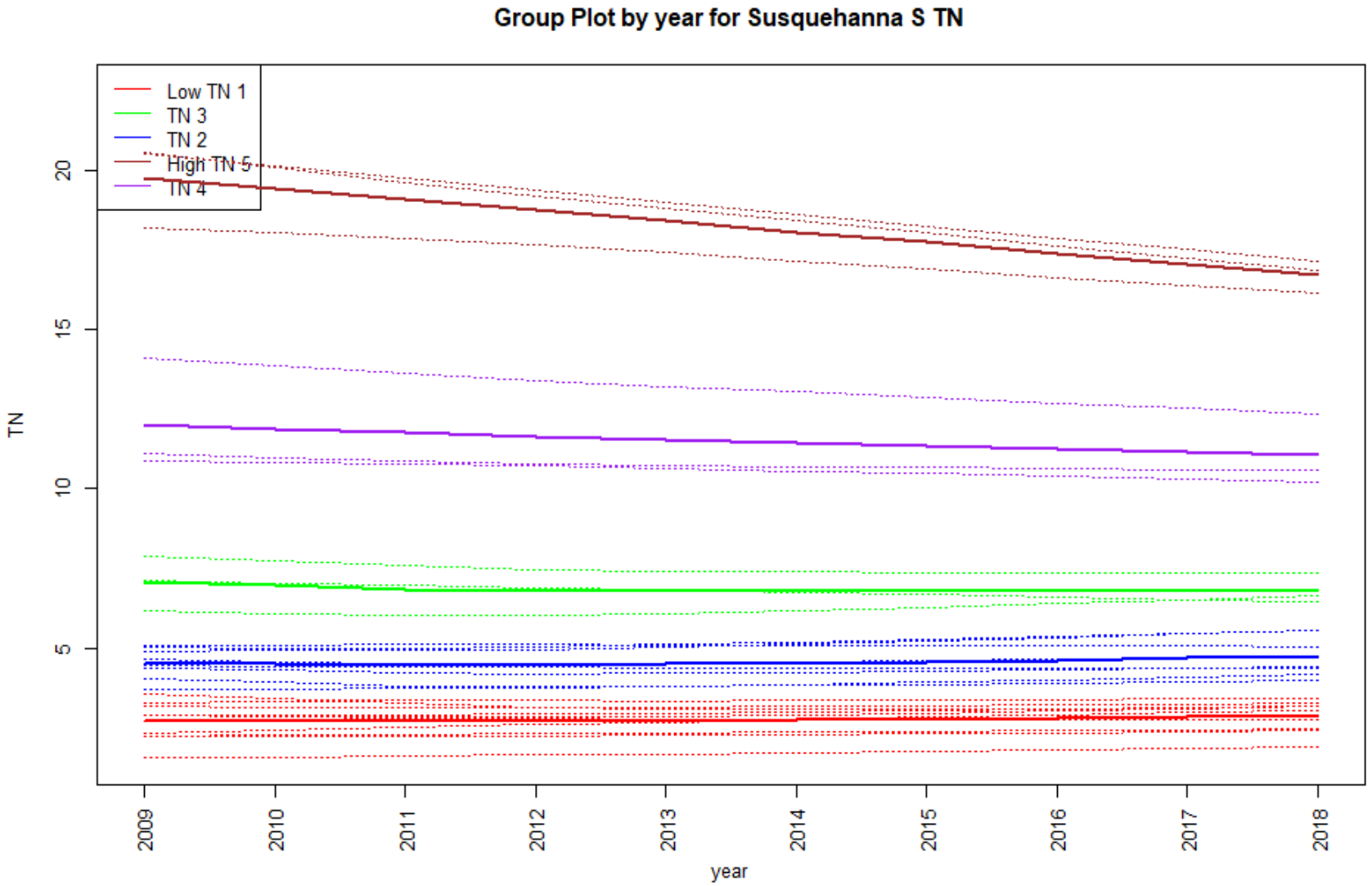
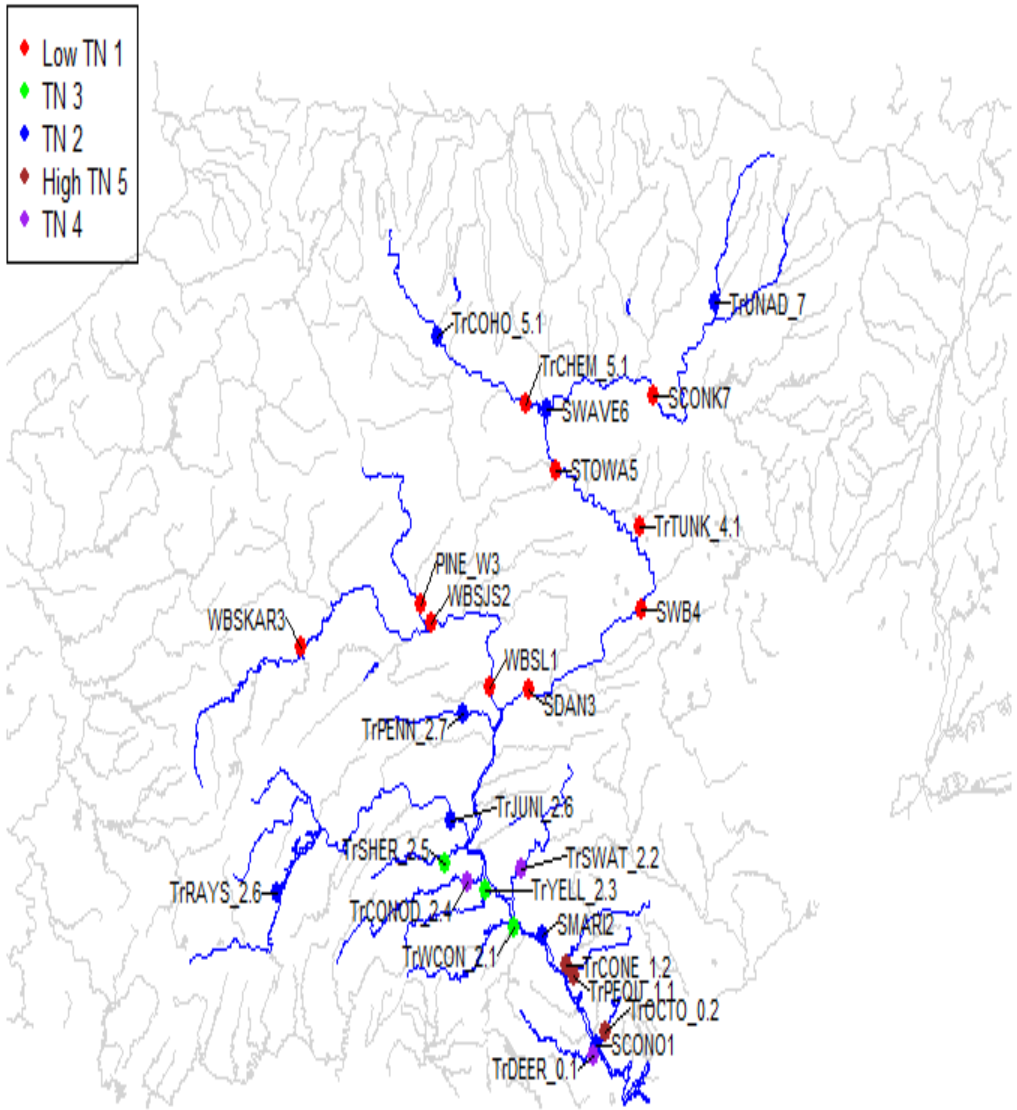


Figure 2.B: Year means plotted with station groups segregated by color. Multiple dash line traces within group show variability among station within groups


Low TN 1
TN 3
TN 2
High TN 5
TN 4
TrCOHO_5.1
TrNAD_7
TrCHEM_5.1
SCONK7
SWAVE6
STOWA5
TrTUNK_4.1
PINE_W3
WBSJS2
WBSKAR3
SWB4
WBSL1
SDAN3
TrPENN_2.7
TrJUNL_2.6
TrSHER_2.5
TrSWAT_2.2
TrRAYS_2.6
TrYELL_2.3
TrCONOD_2.4
SMARI2
TrWCON_2.1
TrCONE_1.2
TrPEQUT_1.1
TrOCTO_0.2
SCONO1
TrDEER_0.1

2. Cluster analysis grouping items defined by station using patterns over year for S TN observed in the Susquehanna. Profiles have been mean adjusted.

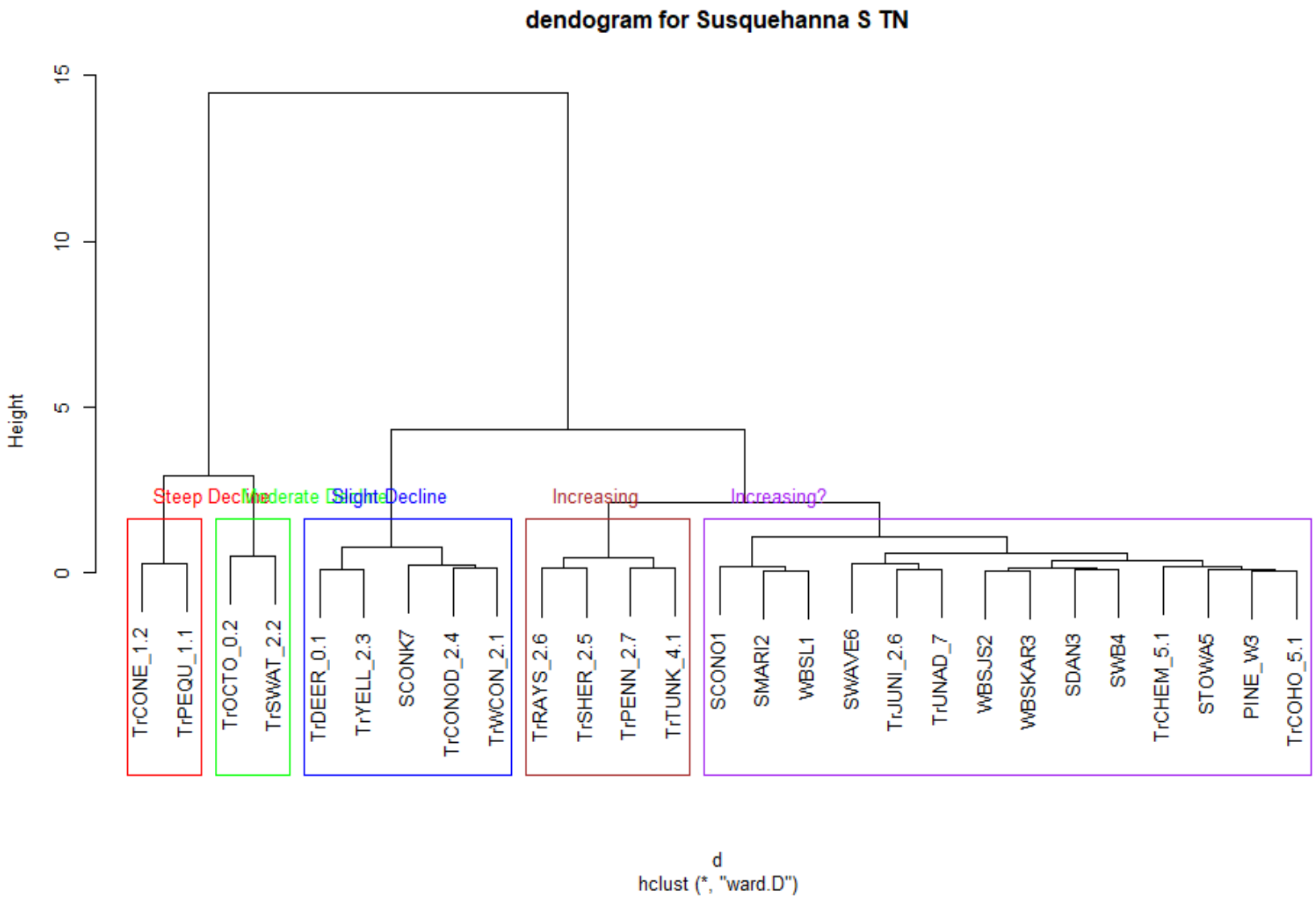


Figure 4.A: Dendogram showing distances between station pairs computed using the dist() function for Susquehanna.

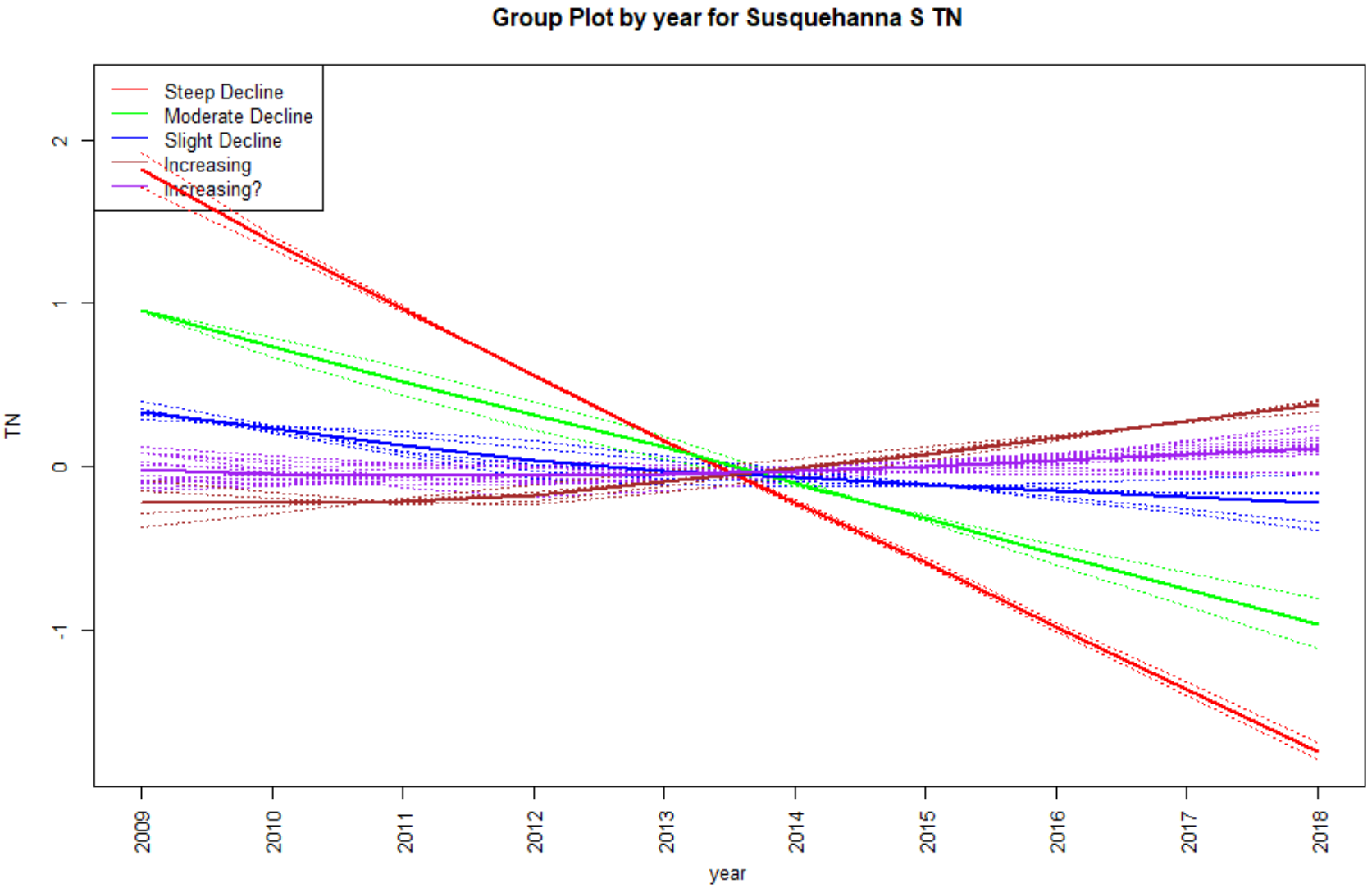
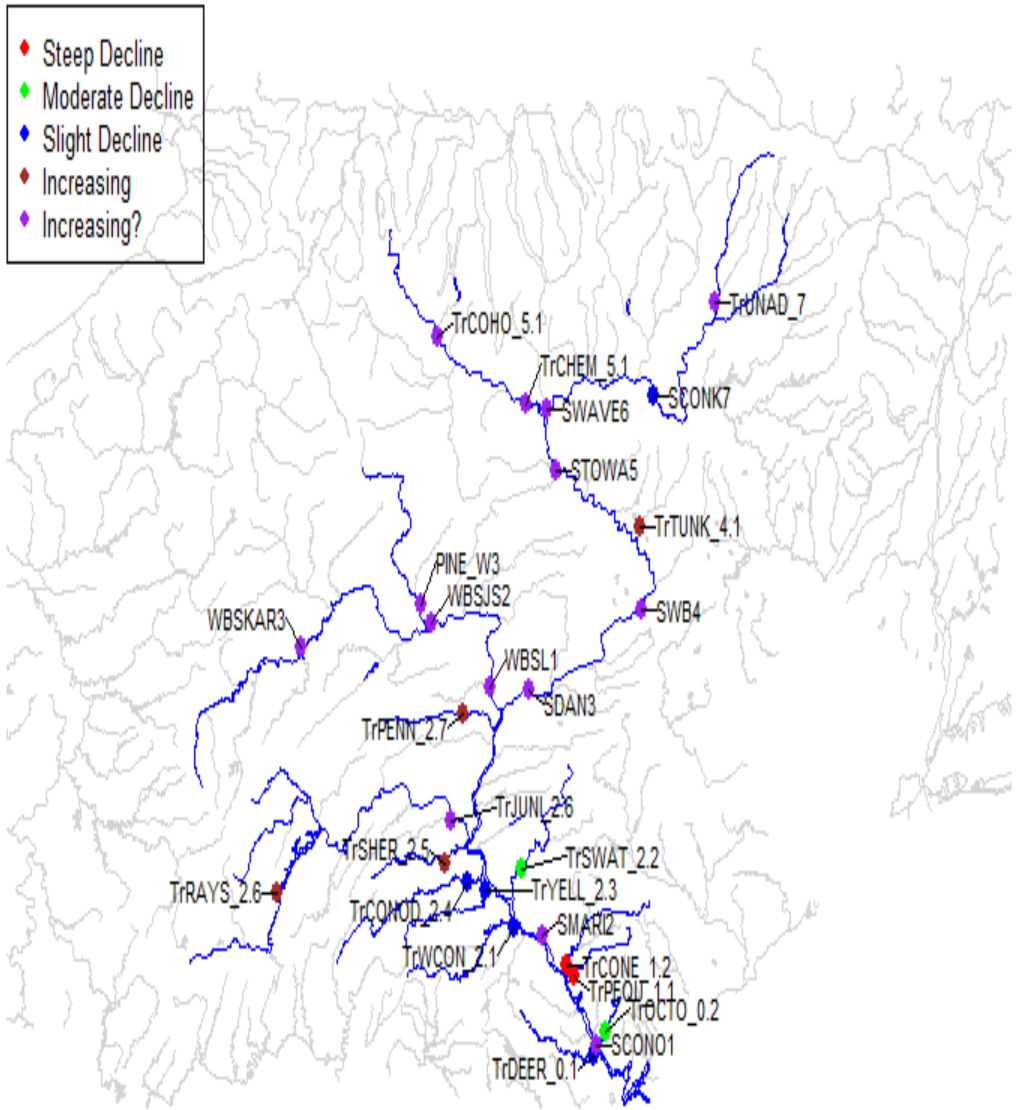


Figure 4.B: Year means plotted with station groups segregated by color. Multiple dash line traces within group show variability among station within groups. Items have been mean adjusted.


Steep Decline
Moderate Decline
Slight Decline
Increasing
Increasing?
TrCOHO_5.1
TrCHEM_5.1
SWAVE6
TrUNAD_7
SCONK7
STOWA5
TrTUNK_4.1
PINE_W3
WBSJS2
WBSKAR3
SWB4
WBSL1
SDAN3
TrPENN_2.7
TrJUNL_2.6
TrSHER_2.5
TrSWAT_2.2
TrRAYS_2.6
TrYELL_2.3
TrCONOD_2.4
SMARI2
TrWCON_2.1
TrCONE_1.2
TrPEQUT_1.1
TrOCTO_0.2
SCONO1
TrDEER_0.1

Crosstabulation of grouping 1and grouping 2.

grp1Lab / grp2Lab	Increasing	Increasing?	Slight Decline	Moderate Decline	Steep Decline
Low TN 1	1	8	1	0	0
TN 2	2	6	0	0	0
TN 3	1	0	2	0	0
TN 4	0	0	2	1	0
High TN 5	0	0	0	1	2

In yellow Low TN sites tend to have increasing trends. In green sites with higher TN tend to be decreasing. In blue sites with low TN and declining trend or high TN and increasing trend are rare.

Comparing Cluster groupings for TN status and TN trend at Susquehanna River Stations.

Station	Status Group	Trend Group
TUNKHANNOCK CREEK NEAR TUNKHANNOCK, PA	Low TN 1	Increasing
PINE CREEK BL L PINE CREEK NEAR WATERVILLE, PA	Low TN 1	Increasing?
SUSQUEHANNA RIVER AT DANVILLE, PA	Low TN 1	Increasing?
SUSQUEHANNA RIVER AT TOWANDA, PA	Low TN 1	Increasing?
SUSQUEHANNA RIVER AT WILKES-BARRE, PA	Low TN 1	Increasing?
CHEMUNG RIVER AT CHEMUNG NY	Low TN 1	Increasing?
WB SUSQUEHANNA RIVER AT JERSEY SHORE, PA	Low TN 1	Increasing?
WB SUSQUEHANNA RIVER AT KARTHAUS, PA	Low TN 1	Increasing?
WEST BRANCH SUSQUEHANNA RIVER AT LEWISBURG, PA	Low TN 1	Increasing?
SUSQUEHANNA RIVER AT CONKLIN NY	Low TN 1	Slight Decline
PENNS CREEK AT PENNS CREEK, PA	TN 2	Increasing
RAYSTOWN BRANCH JUNIATA RIVER AT SAXTON, PA	TN 2	Increasing
SUSQUEHANNA RIVER AT CONOWINGO, MD	TN 2	Increasing?
SUSQUEHANNA RIVER AT MARIETTA, PA	TN 2	Increasing?
SUSQUEHANNA RIVER NEAR WAVERLY NY	TN 2	Increasing?
COHOCTON RIVER NEAR CAMPBELL NY	TN 2	Increasing?
JUNIATA RIVER AT NEWPORT, PA	TN 2	Increasing?
UNADILLA RIVER AT ROCKDALE NY	TN 2	Increasing?
SHERMAN CREEK AT SHERMANS DALE, PA	TN 3	Increasing
WEST CONEWAGO CREEK NEAR MANCHESTER, PA	TN 3	Slight Decline
YELLOW BREECHES CREEK NEAR CAMP HILL, PA	TN 3	Slight Decline
CONODOGUINET CREEK NEAR HOGESTOWN, PA	TN 4	Slight Decline
DEER CREEK NEAR DARLINGTON, MD	TN 4	Slight Decline
SWATARA CREEK NEAR HERSHEY, PA	TN 4	Moderate Decline
OCTORARO CREEK NEAR RICHARDSMERE, MD	High TN 5	Moderate Decline
CONESTOGA RIVER AT CONESTOGA, PA	High TN 5	Steep Decline
PEQUEA CREEK AT MARTIC FORGE, PA	High TN 5	Steep Decline

7. Cluster analysis grouping items defined by year using patterns over station for S TP observed in the Susquehanna Main Branch.

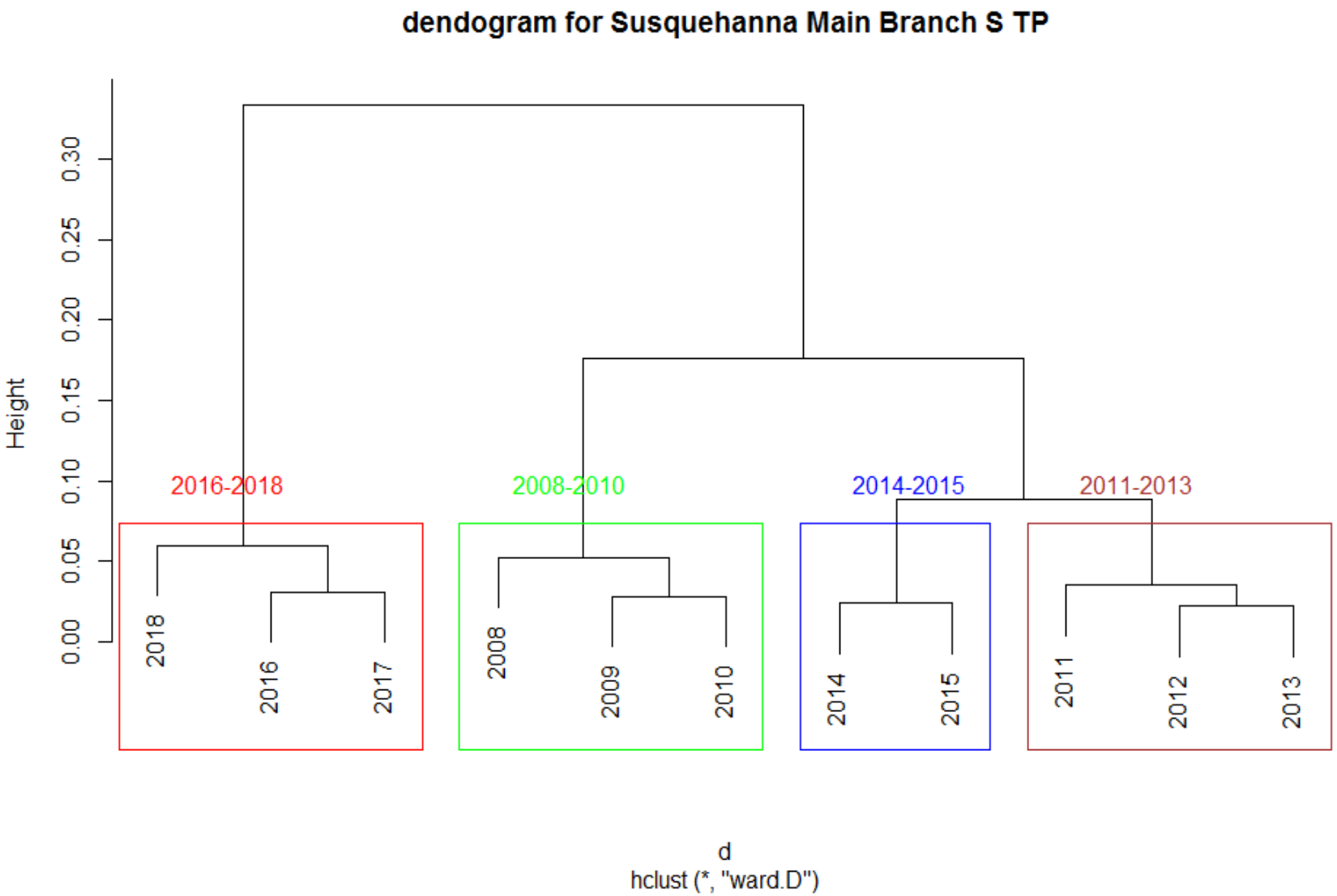


Figure 14.A: Dendogram showing distances between year pairs computed using the dist() function for Susquehanna Main Branch.

Here we examine clusters of years based on the longitudinal profile of the Susquehanna.

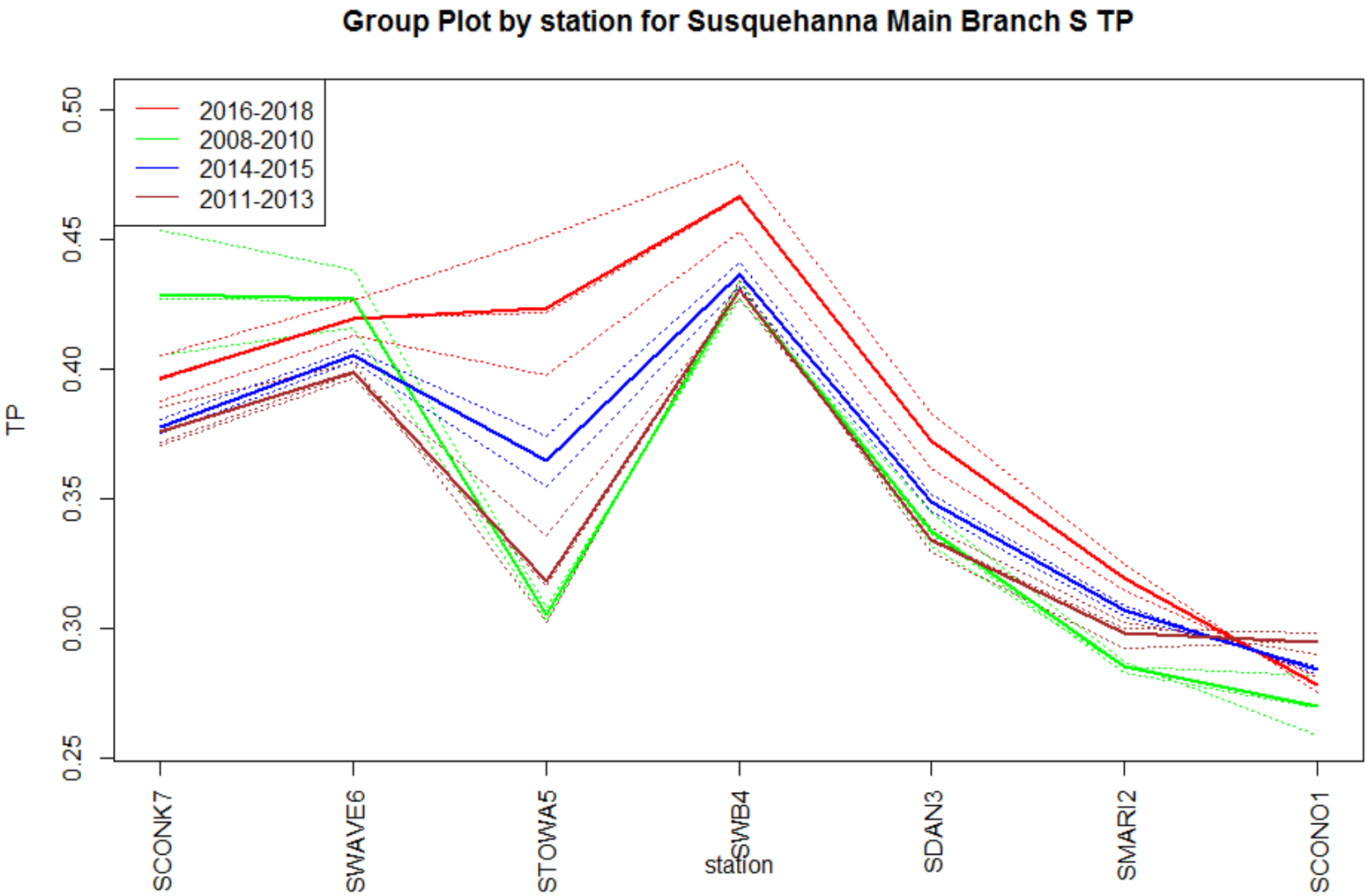


Figure 14.B: Station means plotted with year groups segregated by color. Multiple dash line traces within group show variability among year within groups .

In recent years (red), there is an increasing trend in TP yield from Conklin to Wilkes-barre and then decreasing trend from Wilks-barre to Conowingo. It is interesting that in the early years of this decade (green and brown) Towanda had considerably less TP yield than the upstream (Waverly) and downstream (Wilks-barre) locations. As noted in cluster # 2, Towanda is in a steeply increasing trend group while Waverly is in a flat trend group or possibly slightly U shaped and Wilke-barre is in a slightly increasing trend group. See cluster analysis #2 for more discussion on

this conumdrum.

Another interesting change of long term trend occurs between Marietta and Conowingo. The increasing trend of TP yield established at Towanda continues with some moderation at downstream locations down to Marietta. At Conowingo, which has only a 4.2 % increase in watershed area relative to Marietta, the long term trend pattern changes to have an inverted U shape with a zenith in 2011-2013.

Watershed area ratio Towanda/Waverly (20157.18/12369.81 = 1.629547)
Watershed area ratio Towanda/(Waverly + Chemung) (20157.18/ (12369.8097 + 6508.1475) = 1.067763)
Watershed area ratio Wilkes-barre /Towanda (25795.4598/20157.18 = 1.279716)
Watershed area ratio Danville / Wilkes-barre (29068.8012/25795.4598 = 1.126896)
Watershed area ratio Marietta/Danville (67349.2311/29068.8012 =2.316891)
Watershed area ratio Conowingo/Marietta (70162.0209/67349.2311 =1.041764)

end of file