Modeling Salmon Behavior on the Umpqua River

By Scott Jordan

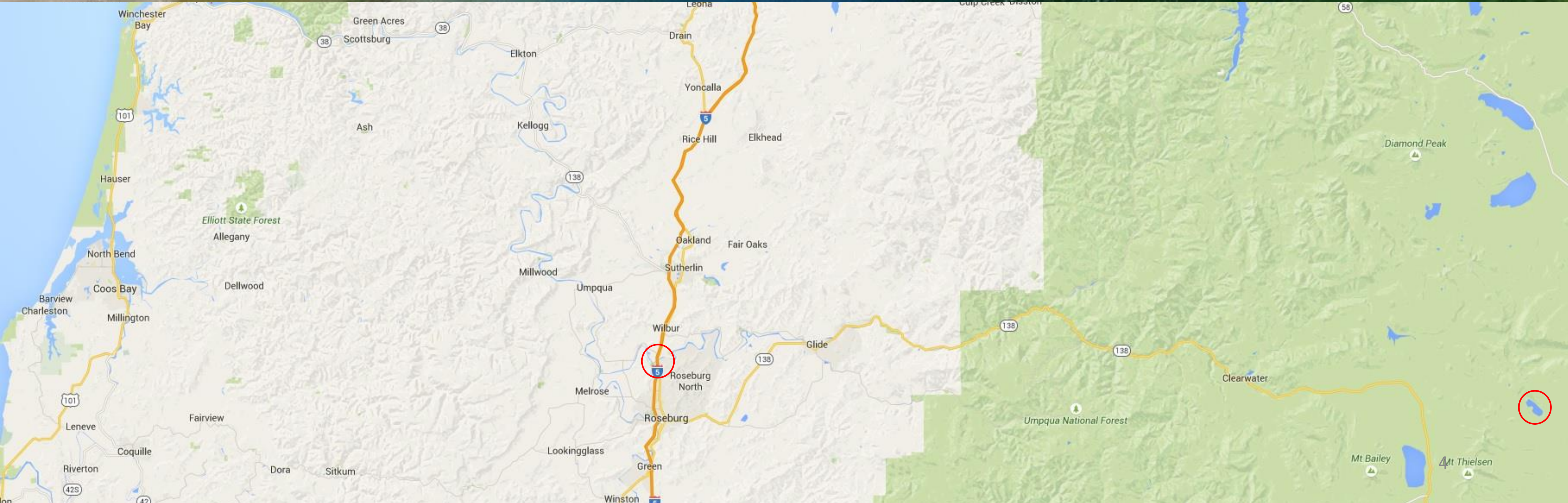
6/2/2015

Importance of Salmon

- Delicious
- Recreation
  - 631,000 people in Oregon went fishing in 2008
  - spent $264.6 Million on fishing trips
- Commercial
  - 2.4 Million pounds of Salmon caught in 2011
  - Catch was worth $6.7 Million
- Conservation
  - Numbers only a fraction of what they once were

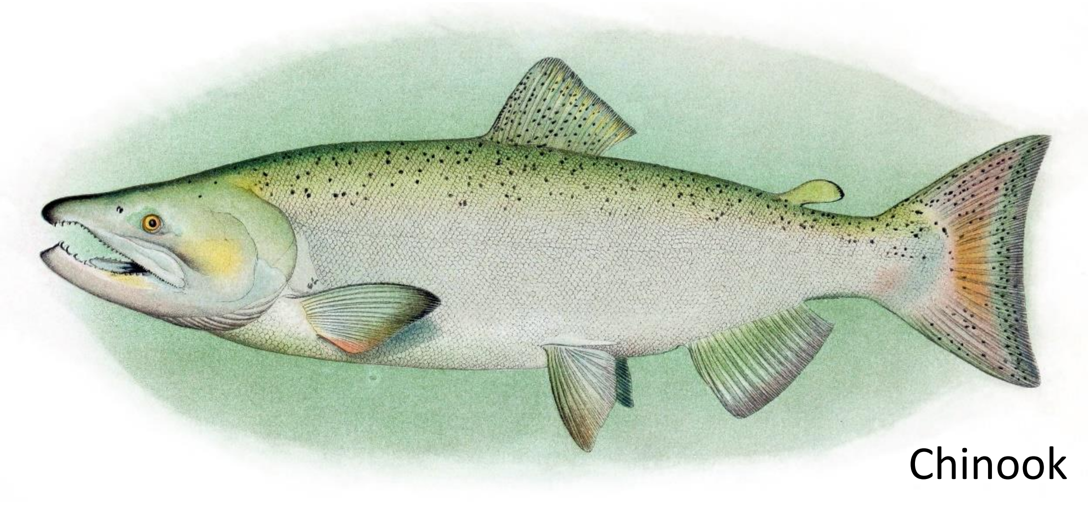
Salmon Migration

- During migration the “run” salmon typically don’t eat, then fertilize and lay eggs, then die
- Vulnerable to predators, dams, fisherman
- Understanding the run can lead to better protection through managing dams and fishing seasons
- Umpqua River is understudied compared to the Columbia

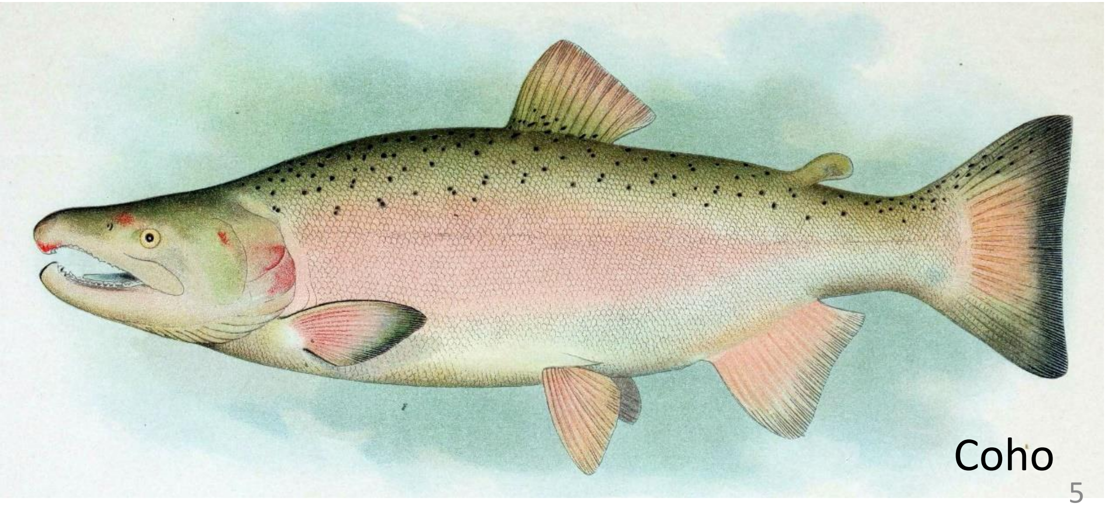

Leona
Winchester Bay
Green Acres
Scottsburg
Drain
Elkton
Yoncalla
Ash
Kellogg
Rice Hill
Elkhead
Diamond Peak
Hauser
Elliott State Forest
Allegany
North Bend
Oakland
Fair Oaks
Sutherlin
Millwood
Umpqua
Coos Bay
Dellwood
Barview
Charleston
Millington
Wilbur
Glide
Clearwater
Roseburg North
Melrose
Roseburg
Umpqua National Forest
Fairview
Leneve
Coquille
Lookingglass
Green
Riverton
Dora
Sitkum
Mt Bailey
Mt Thielsen
Winston

Species that pass Winchester Dam

- Steelhead
- **Chinook**
- Coho
- Brown Trout
- Cutthroat
- Lamprey
- Sockeye
- Chum
- Rainbow
- Sucker



Chinook



Coho

Overall research question

How can archived runs of Chinook Salmon be used to predict future runs?

Goals

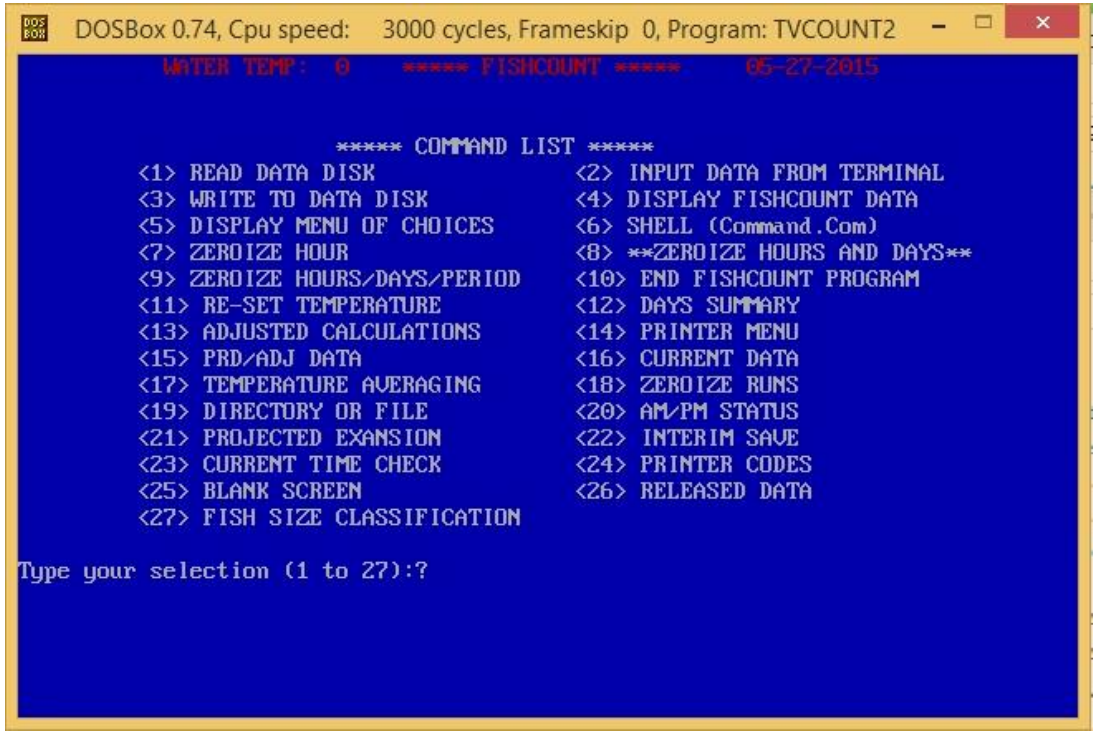
- Data exploration
- Predict running days and non-running days
  - predict the median of the run
  - machine learning algorithms

Expanding the Dataset

- Current dataset had count data from Nov 1998 – Aug 2014
- Archived Data from 1989-1997
  - Camera counting Oct 24, 1991
  - No Description of data file
  - Old MS DOS program used for entry

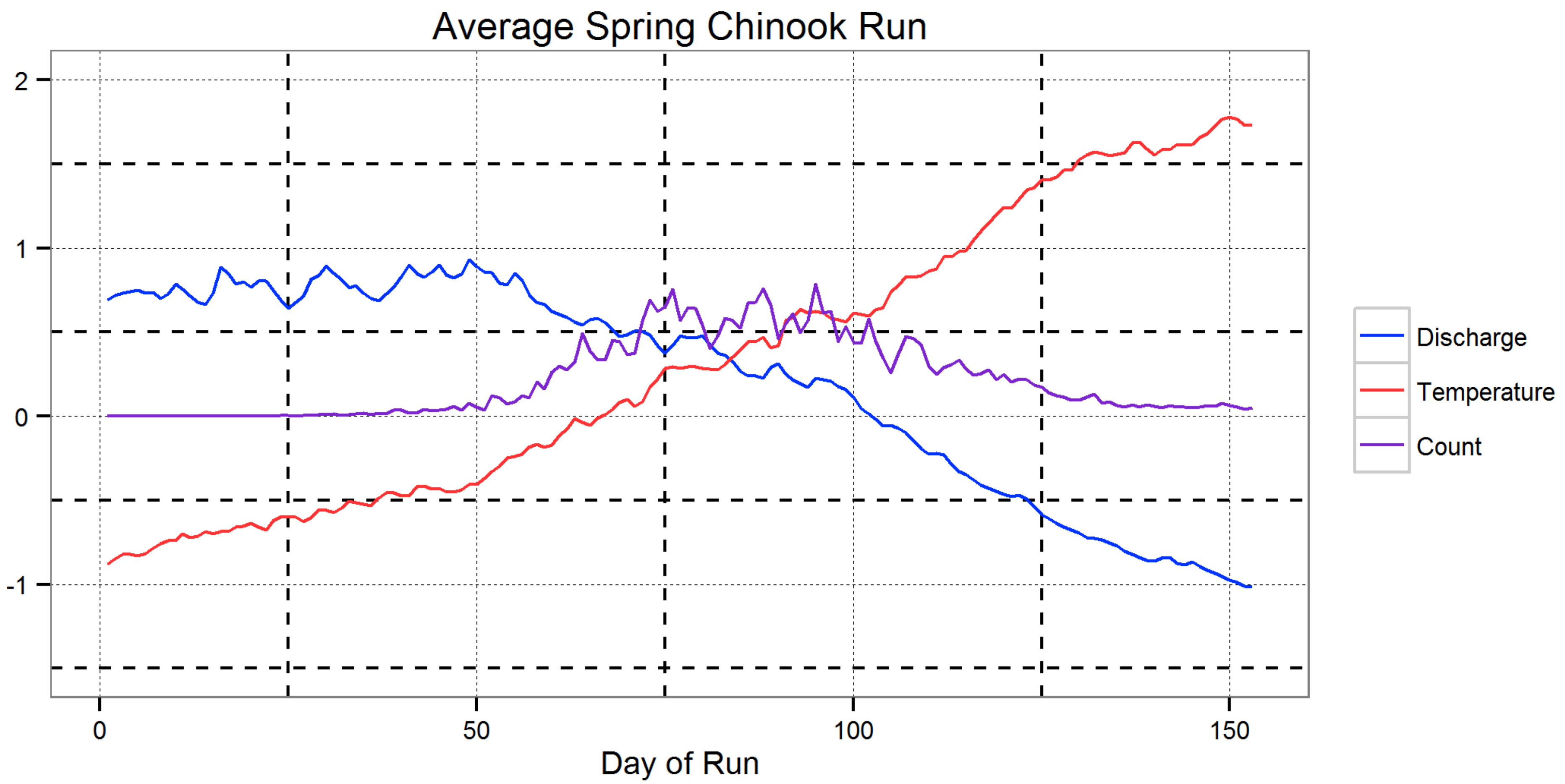
```
0,0,58,1,52
0,0,13,1136
11,11,54,1898
4,4,17,1282
0,0,0,0
0,0,0,11
6,6,11,82
0,0,4,51
0,0,0,0
0,0,0,44
0,0,0,7
0,0,0,0
0,0,3,14
0,0,0,1
0,0,2,702
0,0,0,2
0,0,0,14
9,9,29,58
27,27,81,171
3,3,30,123
0,0,2,11
4,4,13,22
0,0,4,7
1,1,3,6
0,0,0,0
1,1,6,8
0,0,0,0
0,0,0,0
0,0,0,0
0,0,0,0
2,2,8,10
0,0,1,2
0,0,0,0
0,0,0,0
0,0,0,0
0,0,0,0
0,0,0,0
0,0,0,0
0,0,0,0
0,0,0,0
0,0,0,0
0,0,0,0
0,0,0,0
0,0,0,0
0,0,0,0
0,0,14,1680
0,0,74,354
42,42,969,5220
0,0,83,225
0,0,18,64
0,0,0,6
```

Expanding the Dataset

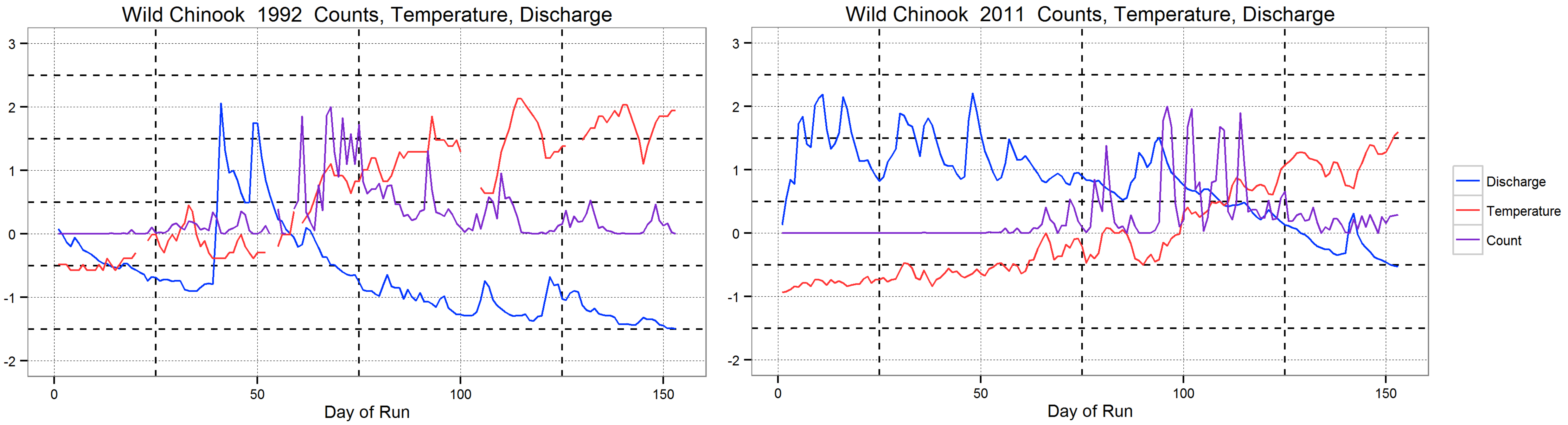


- Reverse engineered the files to identify their structure
- Some data files were missing
- Errors in the data
  - Counts would be wrong
  - Date of the data would be off
  - Fixed most errors through careful analysis
- Total counts for run of Chinook off by small amounts

Spring Chinook



Spring Chinook

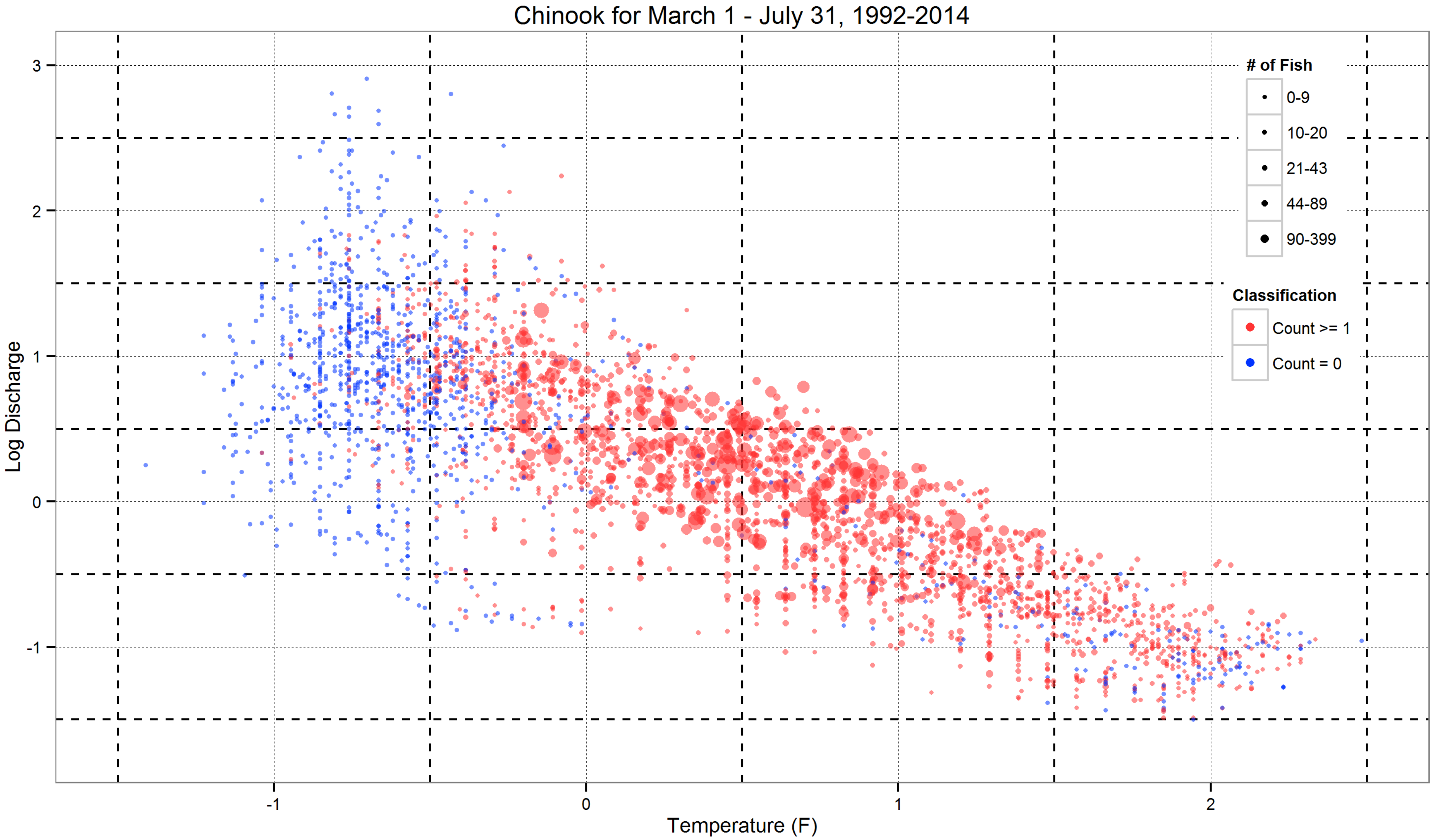


Spring Chinook

Theories of behavior:

- Don’t travel at high level of discharge
- Wait for temperature to rise and river to slow to start run
- Like to travel in groups

What is the Optimal Run Conditions?



Finding Optimal Discharge and Temperature Conditions for Spring Chinook Runs

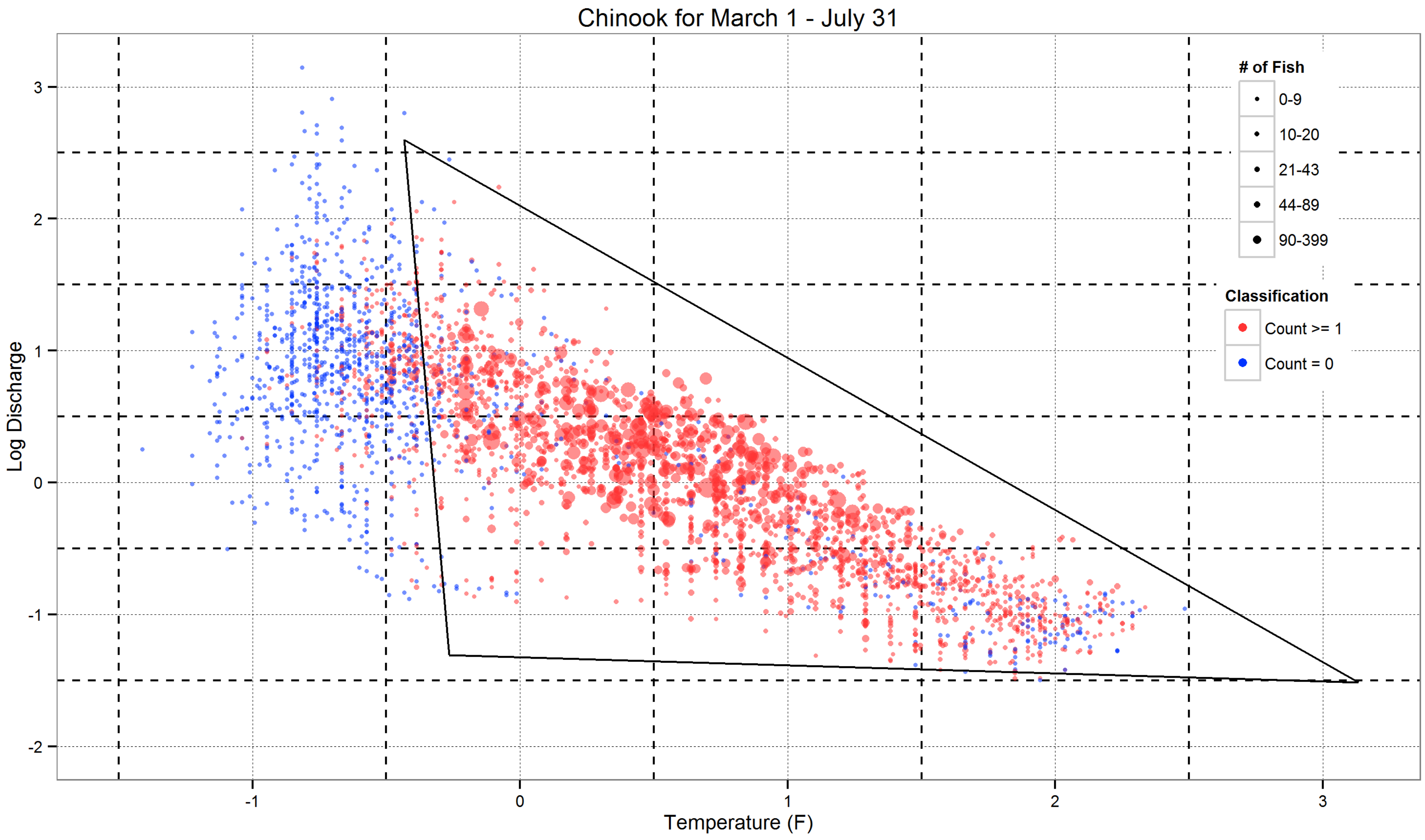
Brute Force

- Exponential problem size
  - Memory
  - Run time
- GPU
  - 70 Million vertex combinations ~ 1 second

Linear/Quadratic Programing

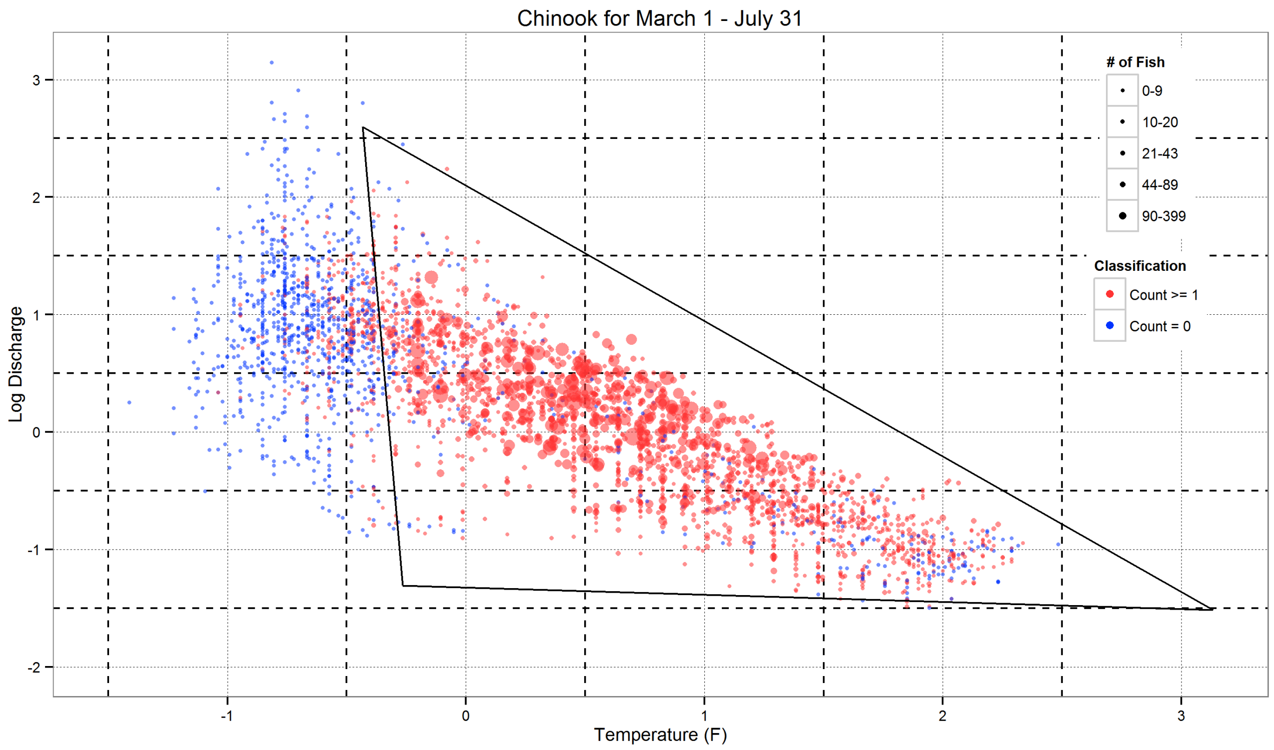
- Write problem as optimization of distances to boundary
- Slow because classifying requires checking each line
- Triangles found on a data set for one year not a representative of the run conditions

Region of Run Conditions



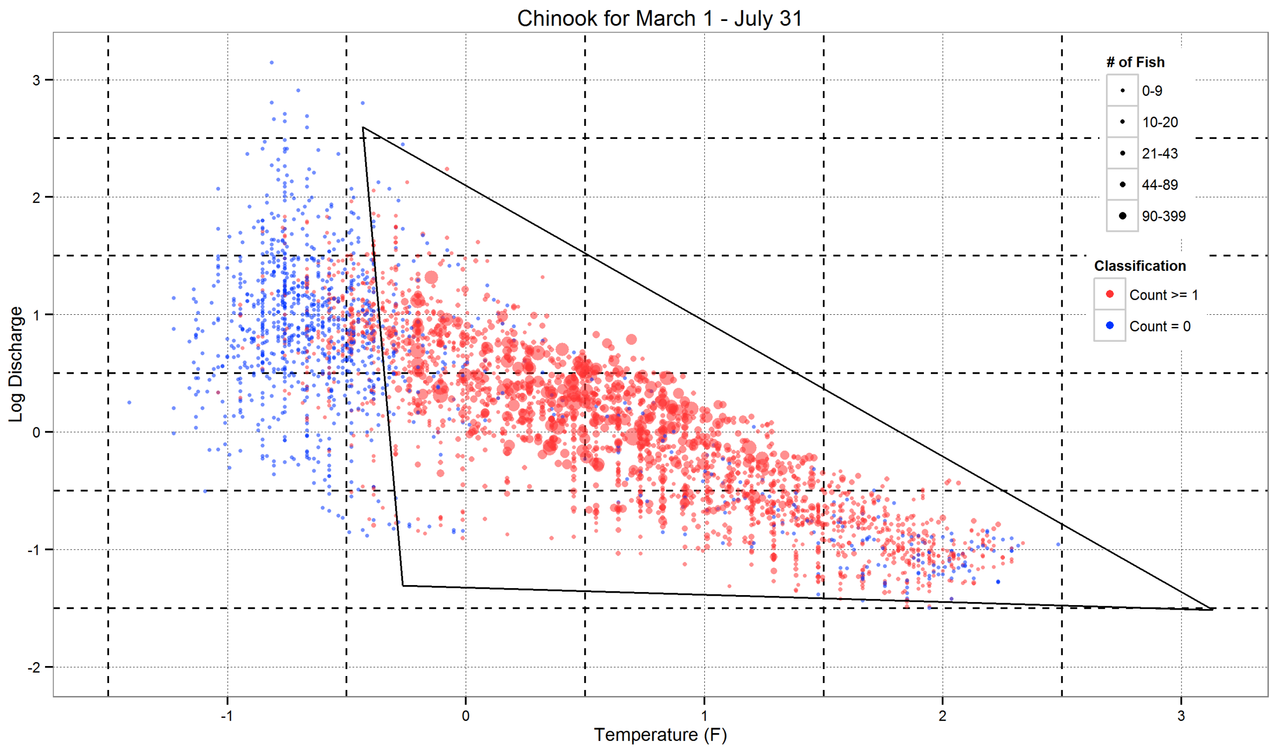
Performance of Region

- Classification Accuracy: 84.09%
- Recalls 90.1% of run days and 72.1% not run days
- Correctly classifies run days 86.5% of the time and not run days 78.5%

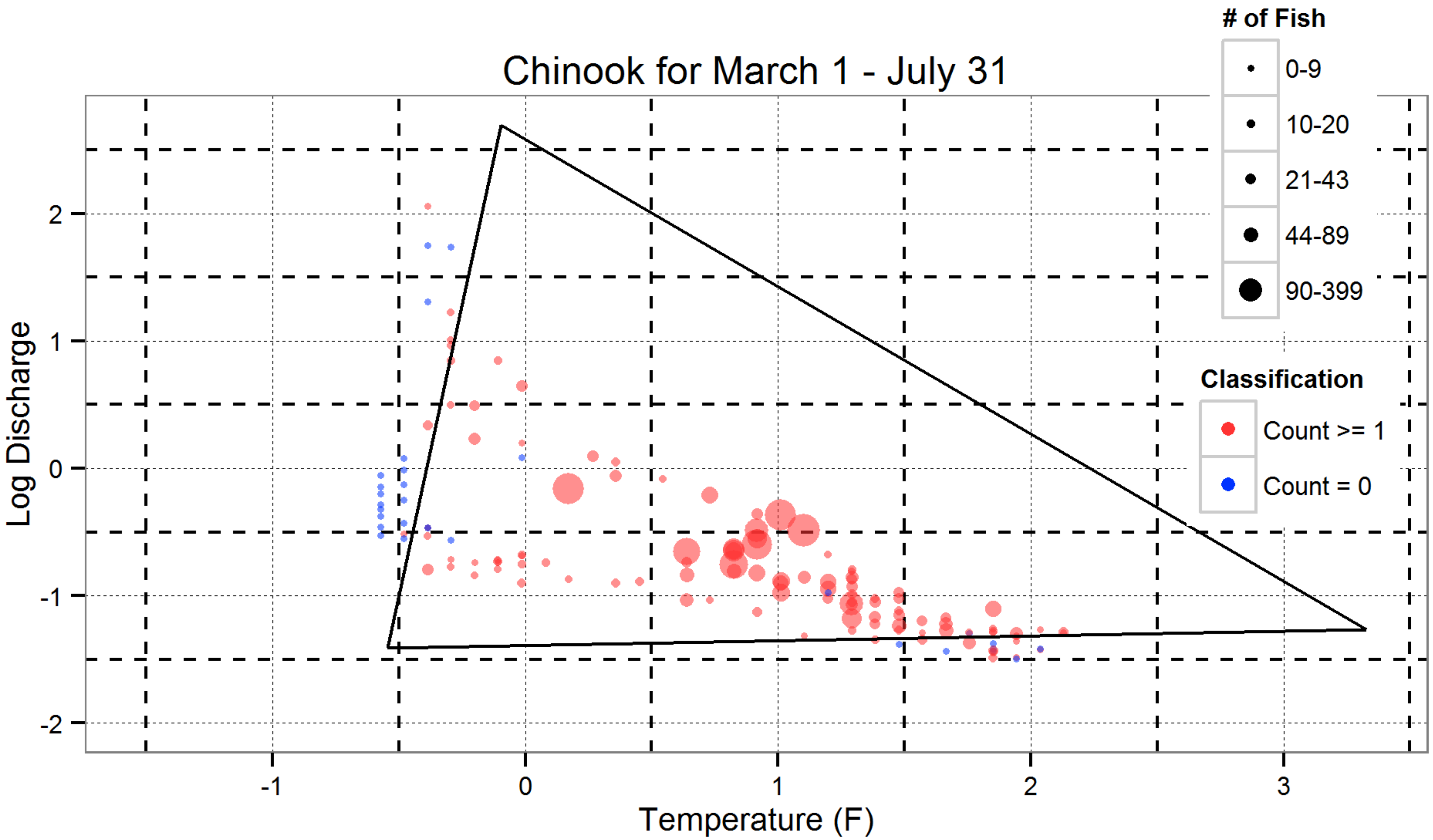


Where the Region Fails

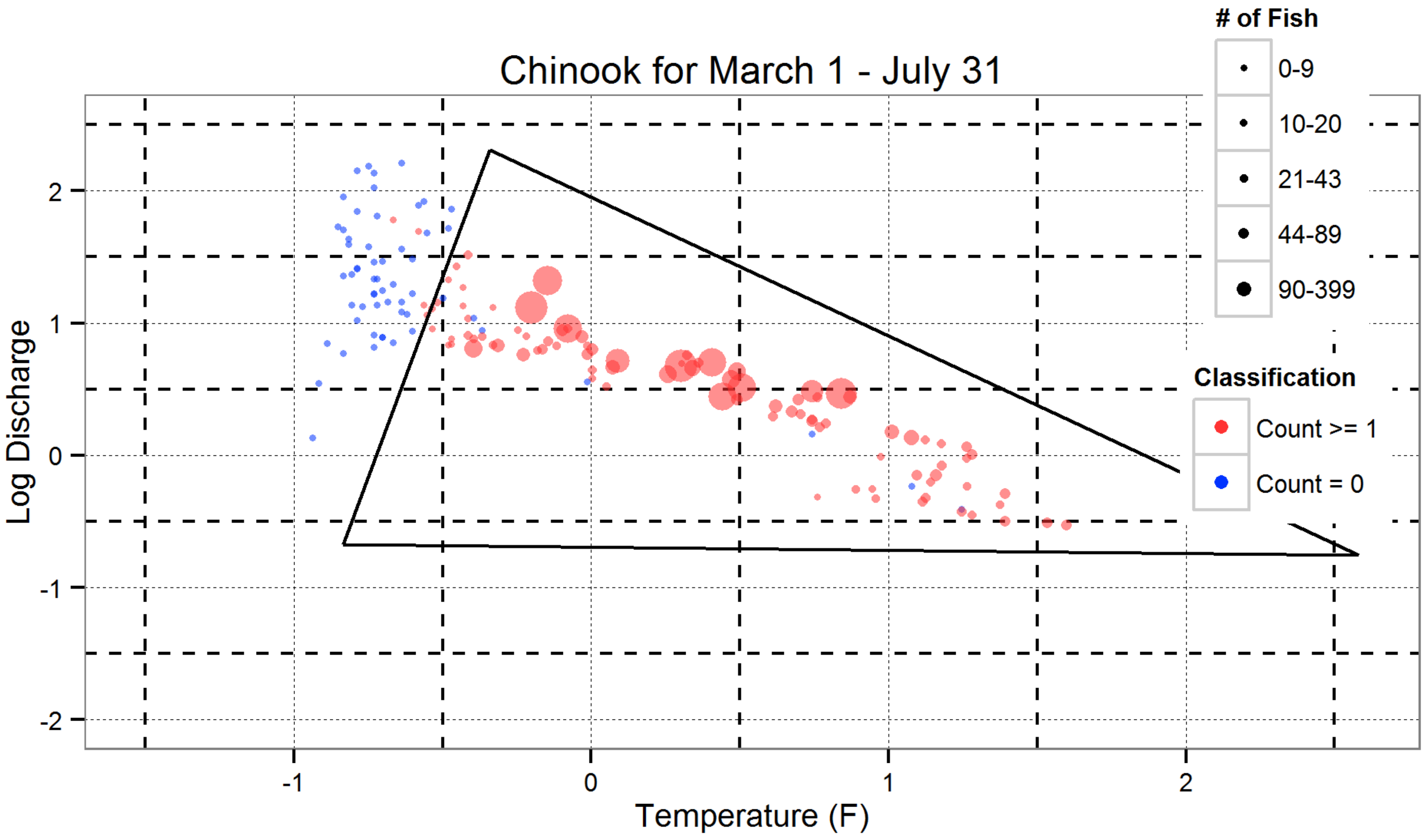
- Misses small run days in beginning
- Includes not run days at the end of the run
- Misses days with in middle of run where no fish come
- Wide variation in run conditions from year to year

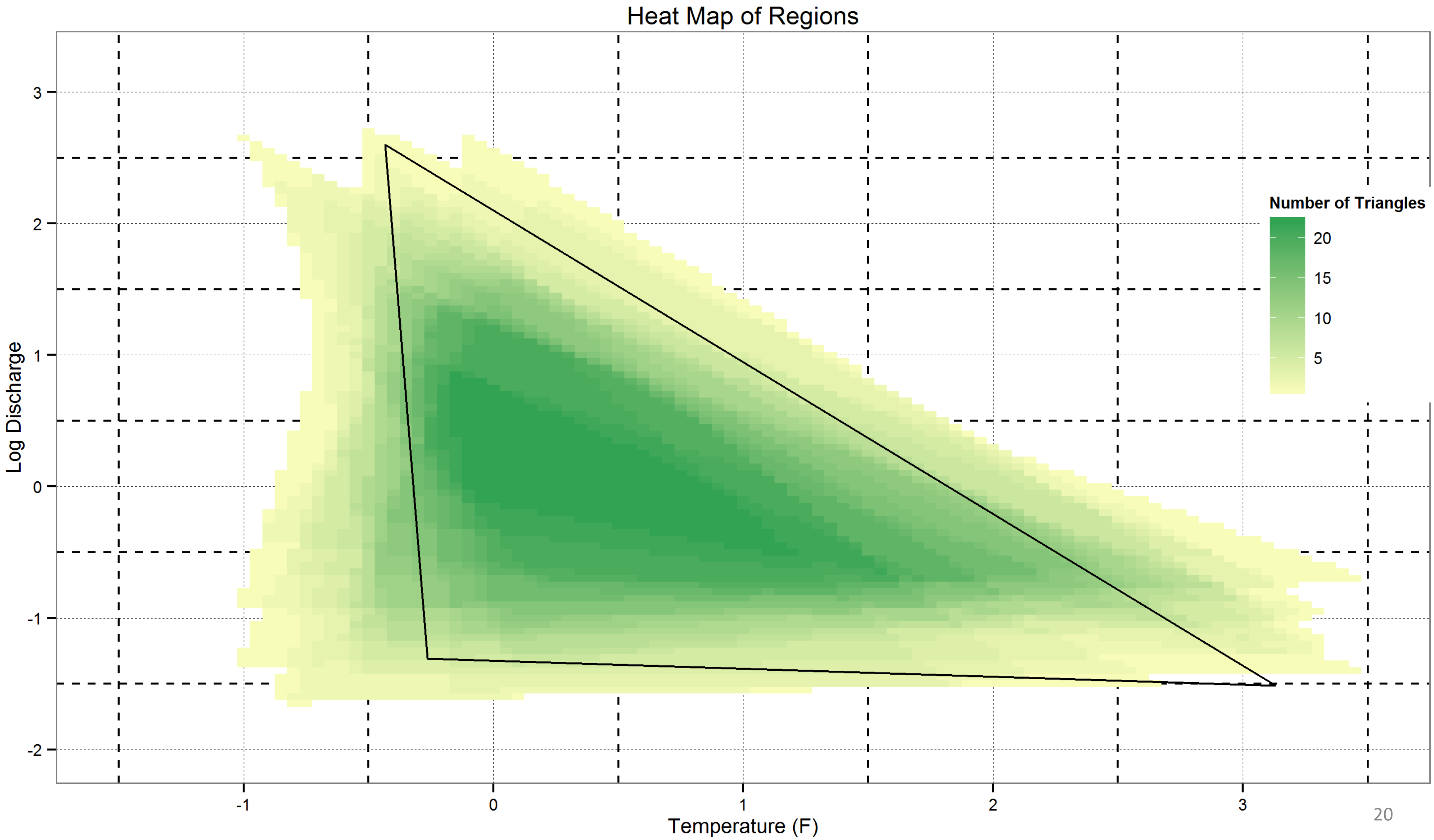


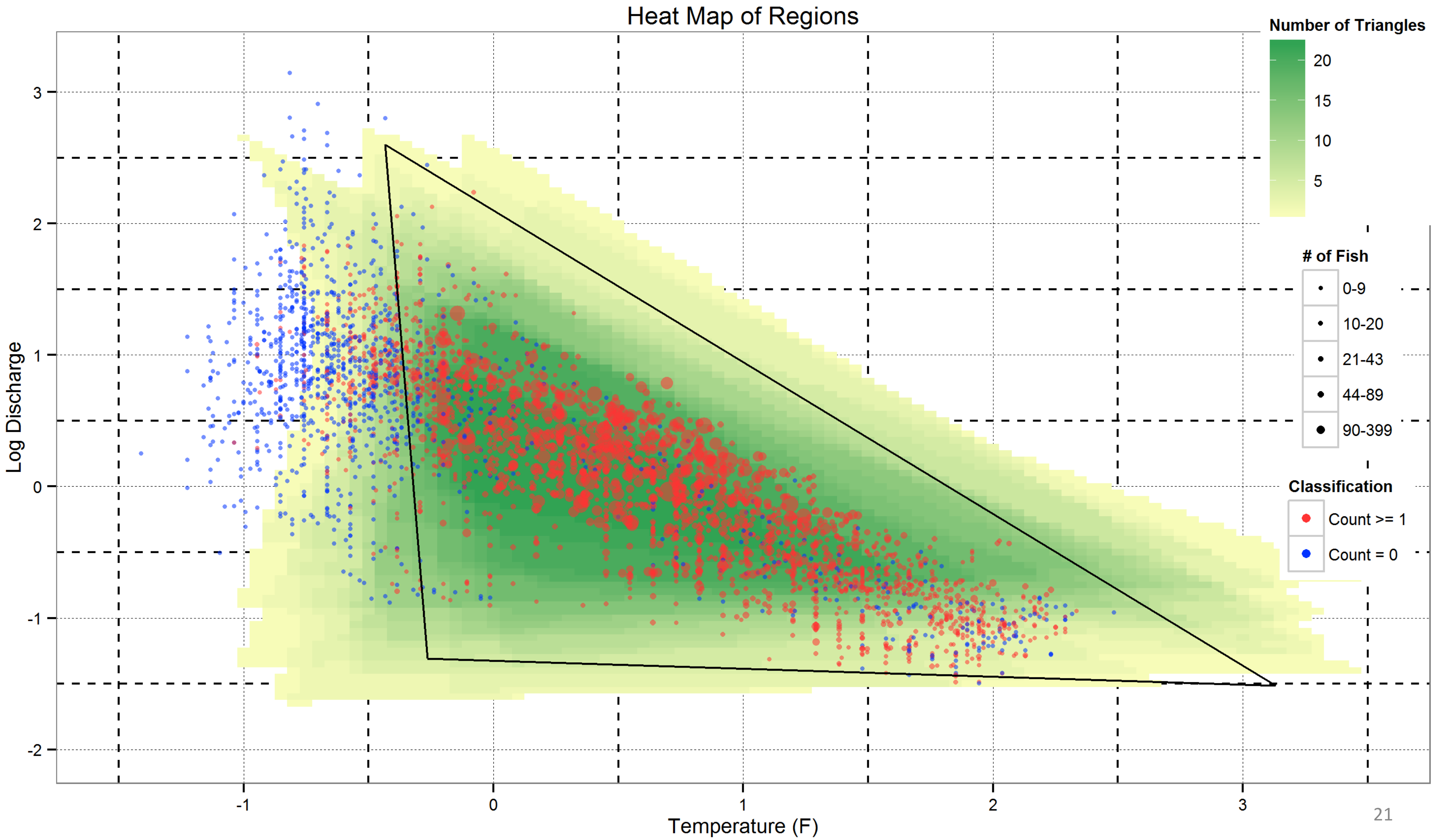
Region for Individual Years - 1992



Region for Individual Years - 2011




Heat Map of Regions
Number of Triangles
20
15
10
5
3
2
1
0
-1
-2
Log Discharge
-1
0
1
2
3
Temperature (F)


Heat Map of Regions
Number of Triangles
20
15
10
5
# of Fish
0-9
10-20
21-43
44-89
90-399
Classification
Count >= 1
Count = 0
Log Discharge
3
2
1
0
-1
-2
Temperature (F)
-1
0
1
2
3

Data exploration: Conclusions

Found a region for the main portion of the run

- Temperature Range Roughly: 55.1˚F to 73.2˚F
- Discharge Range Roughly: 1850 ft^3/sec to 7855 ft^3/sec

Lots of year to year variation in water temperature and discharge results in “cloudy” areas near the edge of the region

Goals

- Data exploration
- Prediction running days and non-running days

  - predict the median of the run
  - machine learning technique

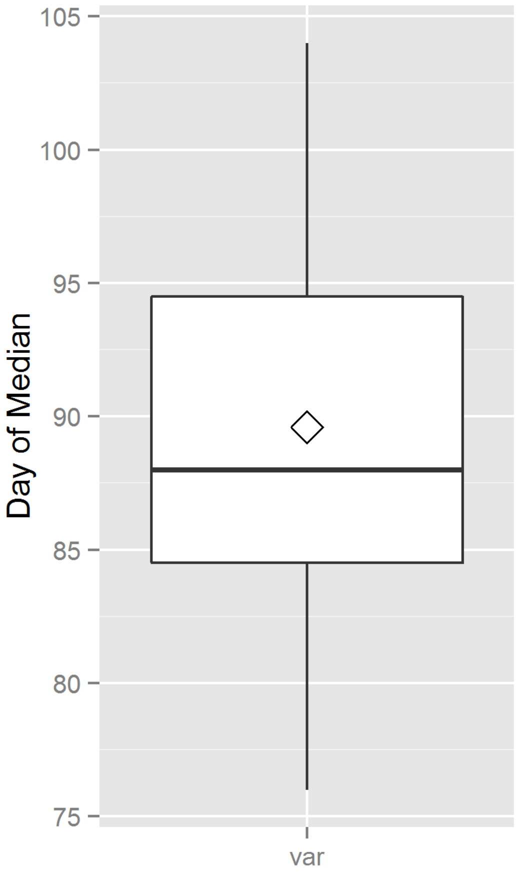
Predicting Median Run

Keefer et al. (2008) Migration Timing of Columbia River Spring Chinook Salmon: Effect of Temperature, River Discharge, and Ocean Environment

- Monthly Variables for January-April:
  - Discharge
  - Air temperature
  - Pacific Decadal Oscillation (PDO)
  - North Pacific Index (NPI),
- Best model: April flow + Jan NPI + Jan PDO with r^2 of .49

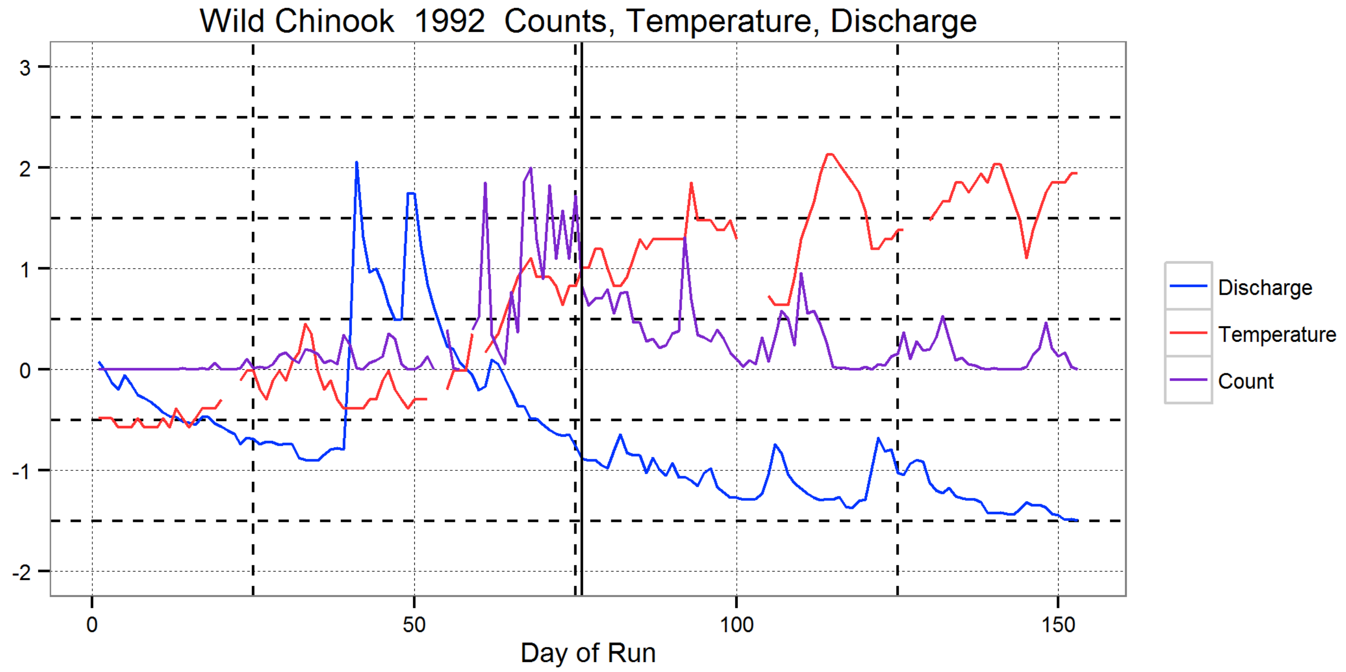
Predicting the Median of the Run

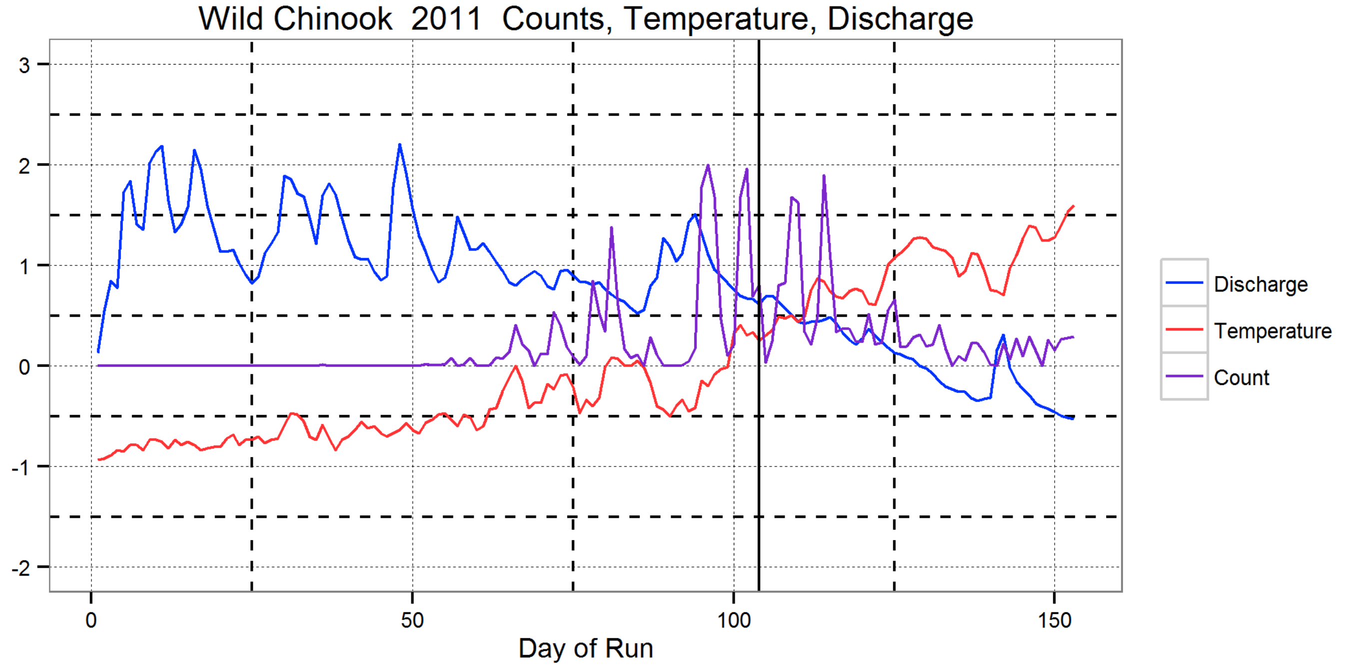
- Earliest Median Day:
  May 15^{th}, 1992
- Latest Median Day:
  June 12^{th}, 2011
- Average Median Day:
  - May $28/29^{th}$
- Median, Median Day:
  - May 27th

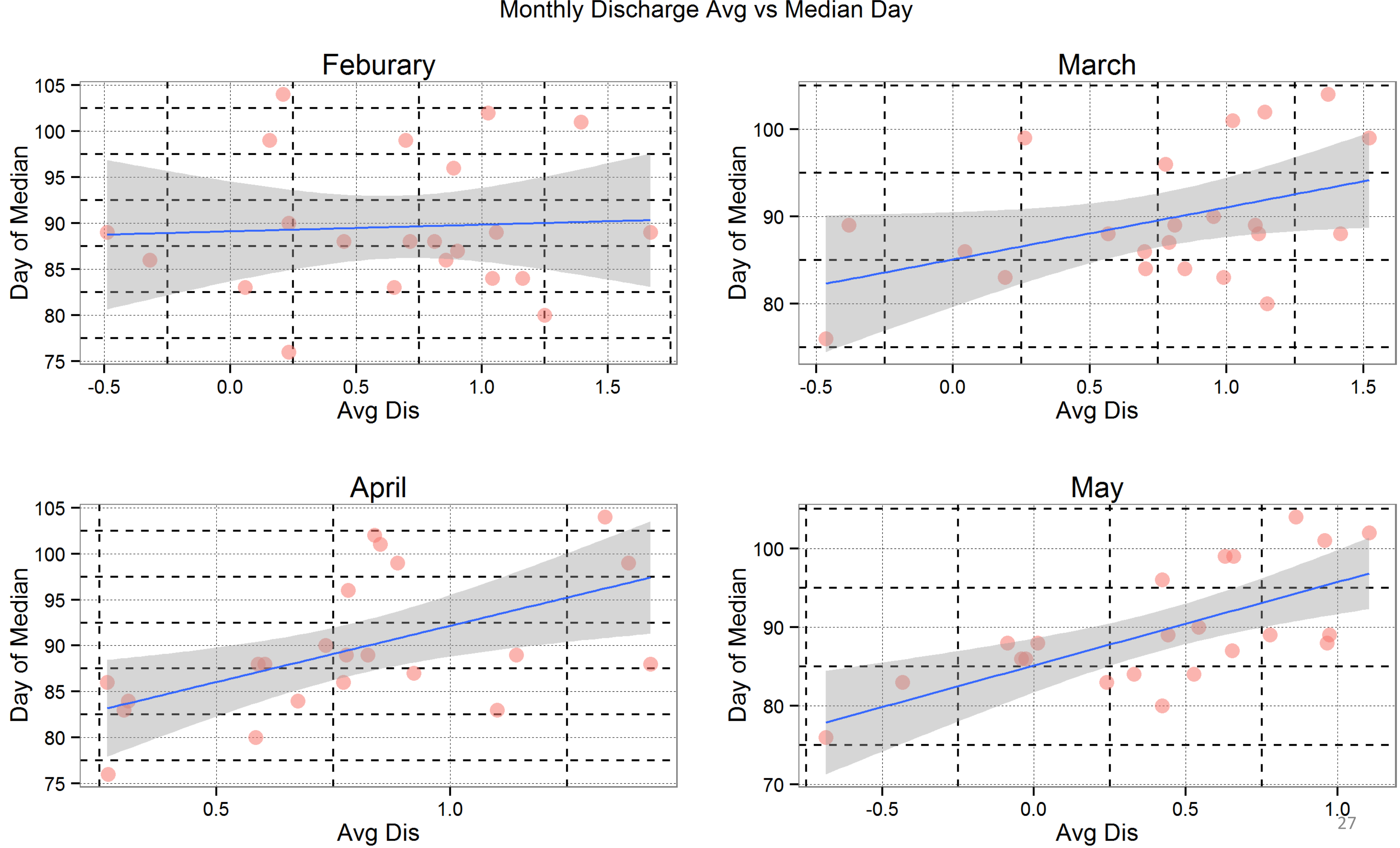


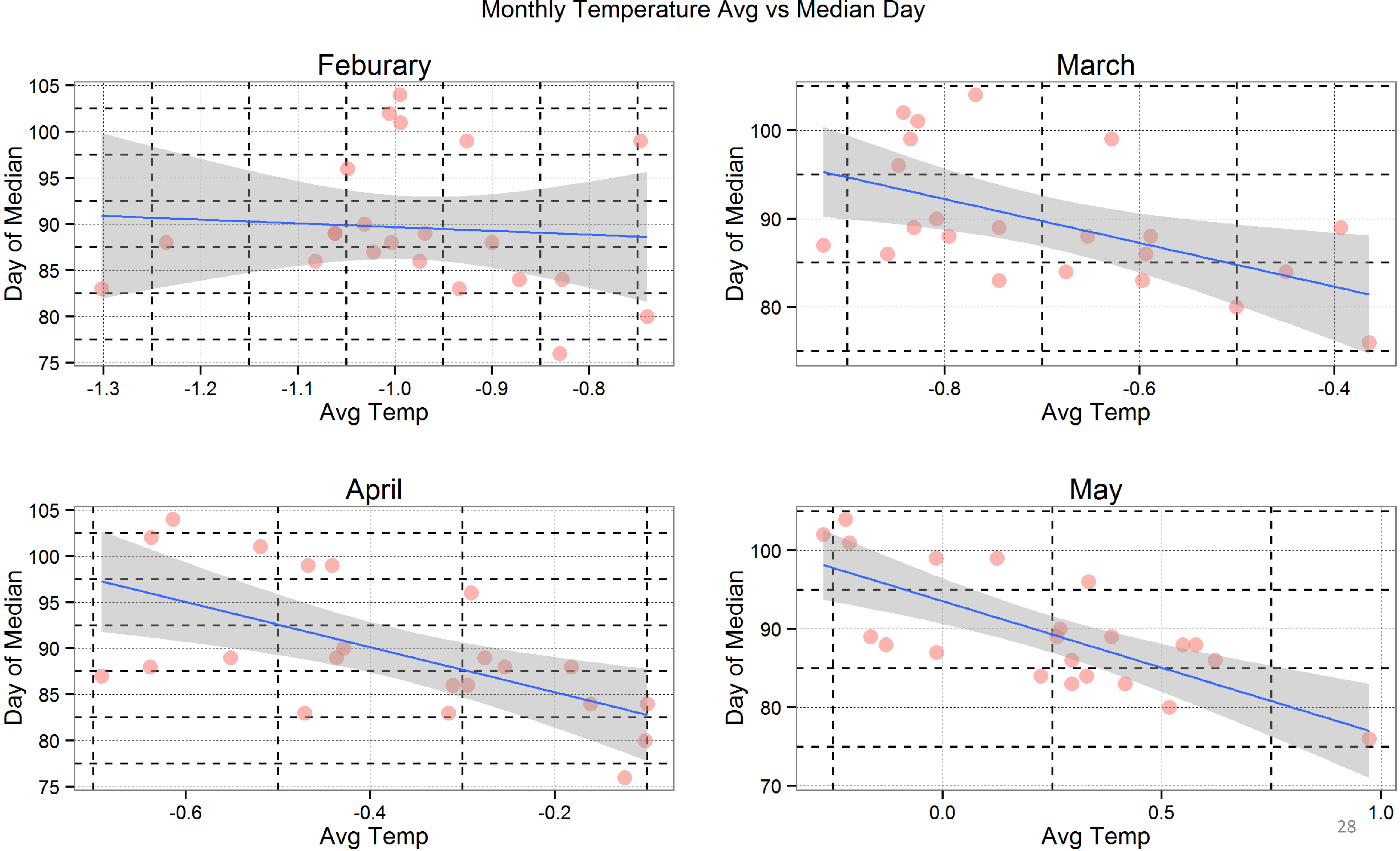
Predicting the Median of the Run

- Low discharge and early warm temperatures mean early run
- Run can be delayed if discharge rises late
- Later runs have higher levels of discharge to start and the river doesn't warm up until late
- Maybe snow melt is the cause of late high discharge

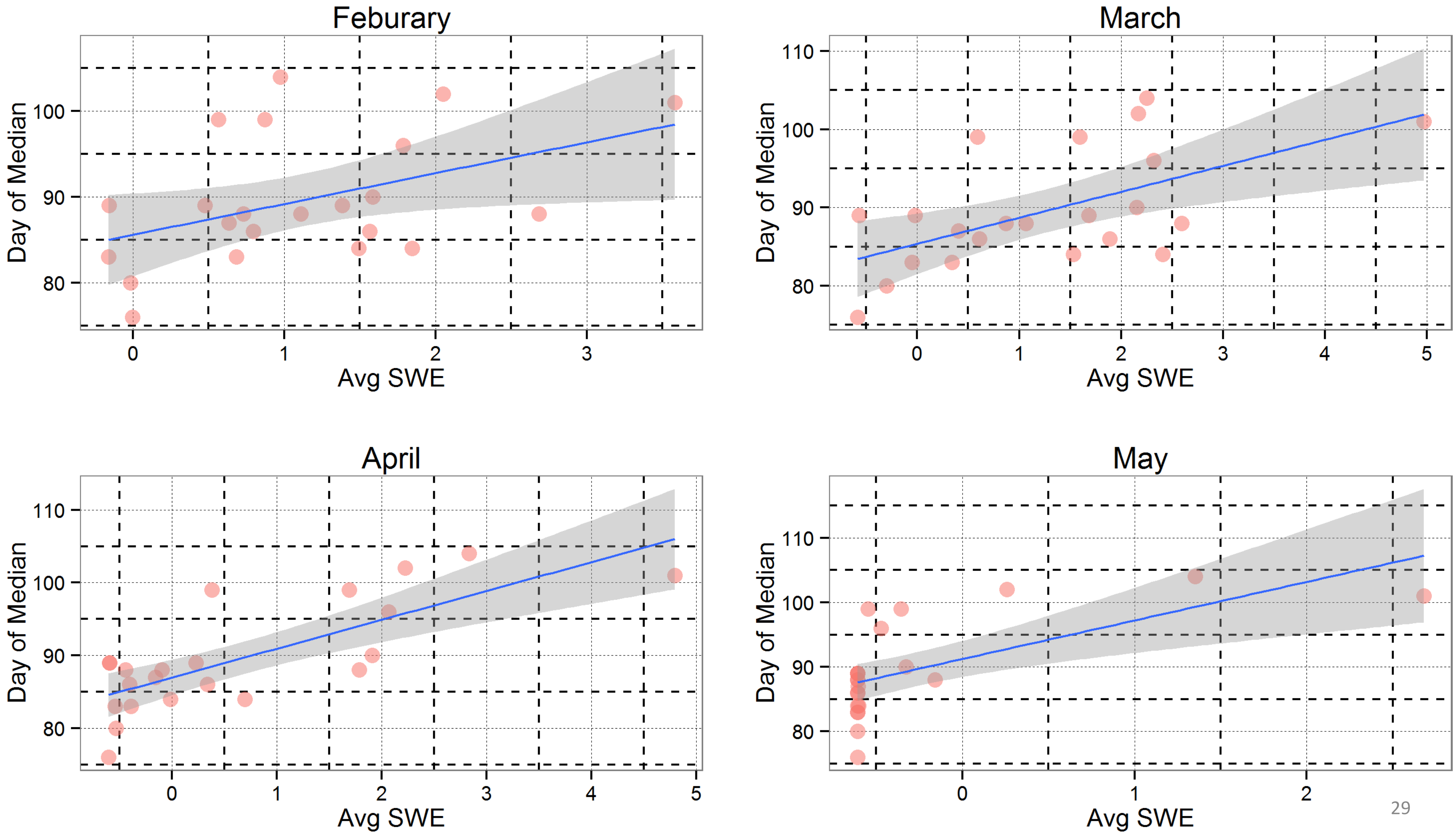


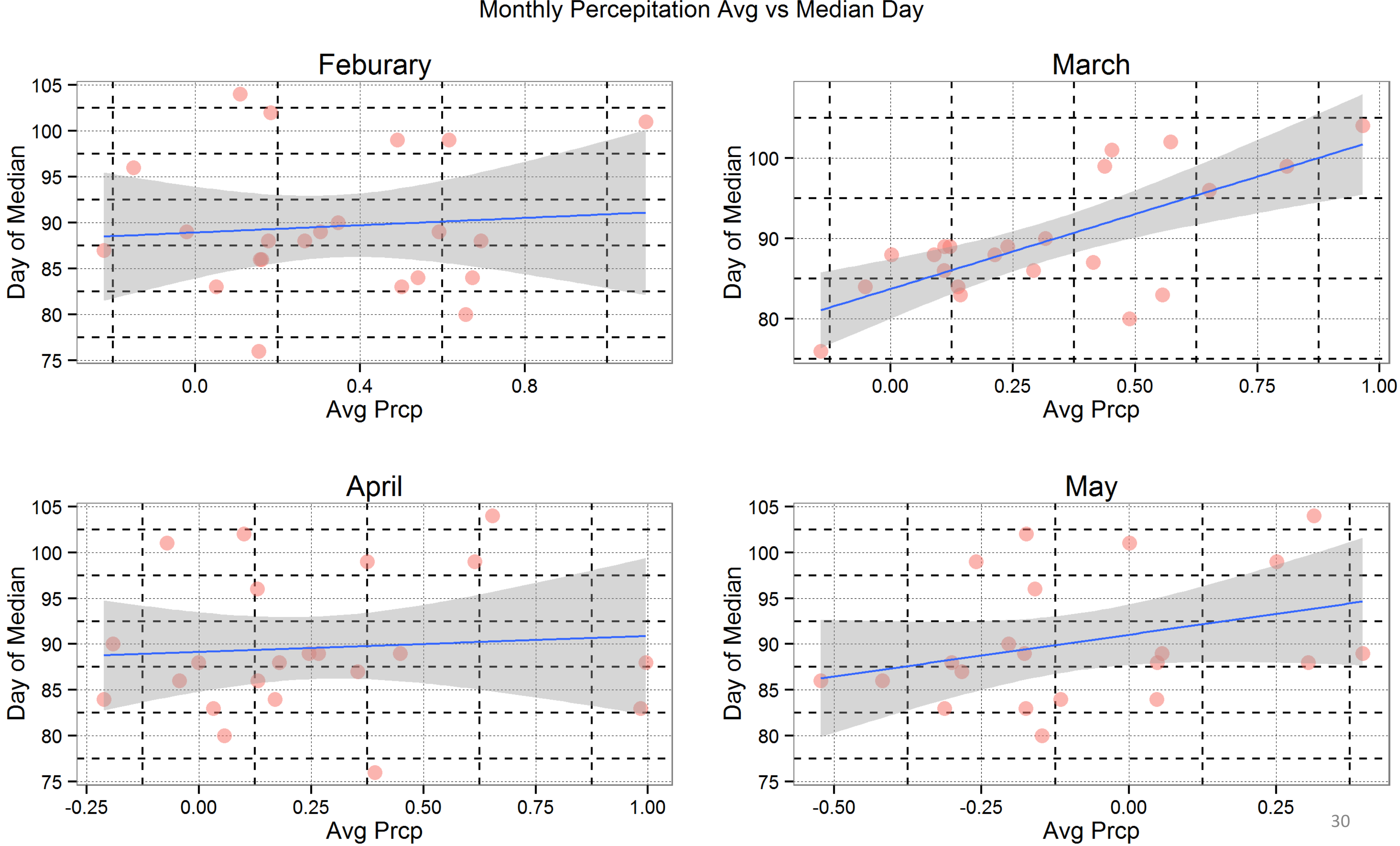



Monthly Discharge Avg vs Median Day
Feburary
March
April
May
Day of Median
Avg Dis
105
100
95
90
85
80
75
70
-0.5
0.0
0.5
1.0
1.5


Monthly Temperature Avg vs Median Day
Feburary
March
April
May
Day of Median
Avg Temp
105
100
95
90
85
80
75
70
-1.3
-1.2
-1.1
-1.0
-0.9
-0.8
-0.6
-0.4
-0.2
0.0
0.5
1.0

Monthly Snow Water Equivalent (SWE) Avg vs Median Day

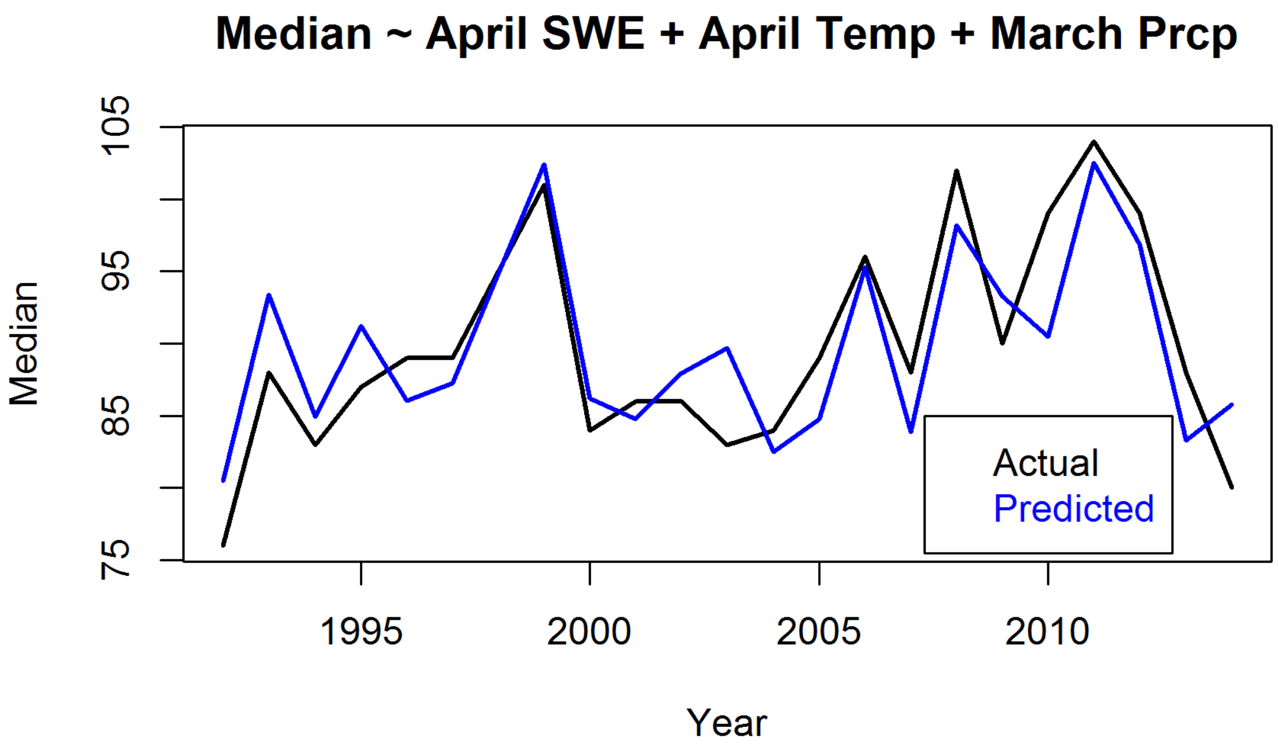


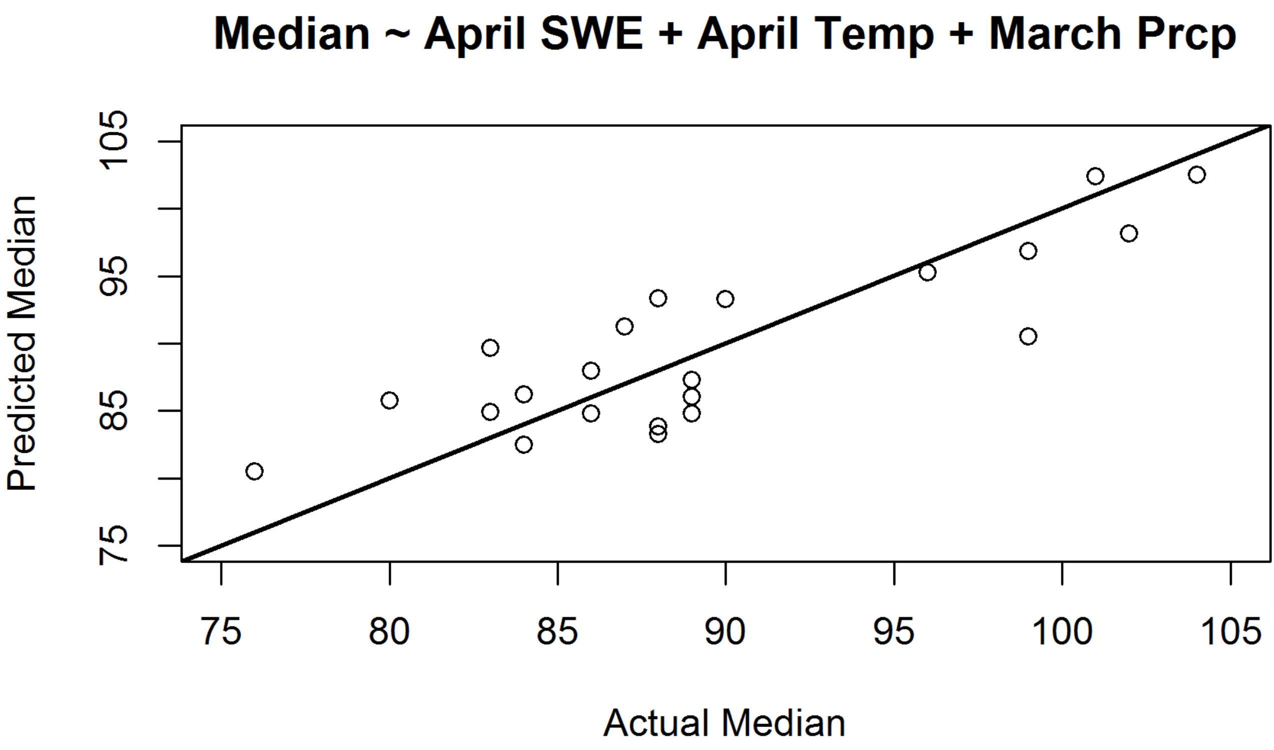


Monthly Percepitation Avg vs Median Day
Feburary
March
April
May
Day of Median
Avg Prcp
105
100
95
90
85
80
75
0.0
0.4
0.8
0.00
0.25
0.50
0.75
1.00
-0.25
-0.50

Model	r^2	P-Val	Sd Error
Temp March	0.2858	0.0103	6.32
Temp April	0.3639	0.0029	5.97
Dis March	0.1831	0.0469	6.76
Dis April	0.3124	0.0069	6.21
SWE Feb	0.2004	0.0367	6.695
SWE March	0.3427	0.0042	6.07
SWE April	**0.5778**	**4.05E-05**	**4.87**
Prcp March	**0.4918**	**0.0003**	**5.34**
Temp+Dis+SWE April	0.6614	4.20E-06	4.36
Temp+SWE April	0.6540	5.24E-06	4.40
SWE April + Prcp March	**0.6842**	**2.06E-06**	**4.21**
SWE April + Temp April + Prcp March	**0.7134**	**7.68E-07**	**4.01**
SWE April + Temp April + Dis April + Prcp March	0.7139	7.52E-07	4.00

Predicting the Median of the Run





Goals

- Data exploration
- Predict running days and non-running days

 - predict the median of the run

 - machine learning technique

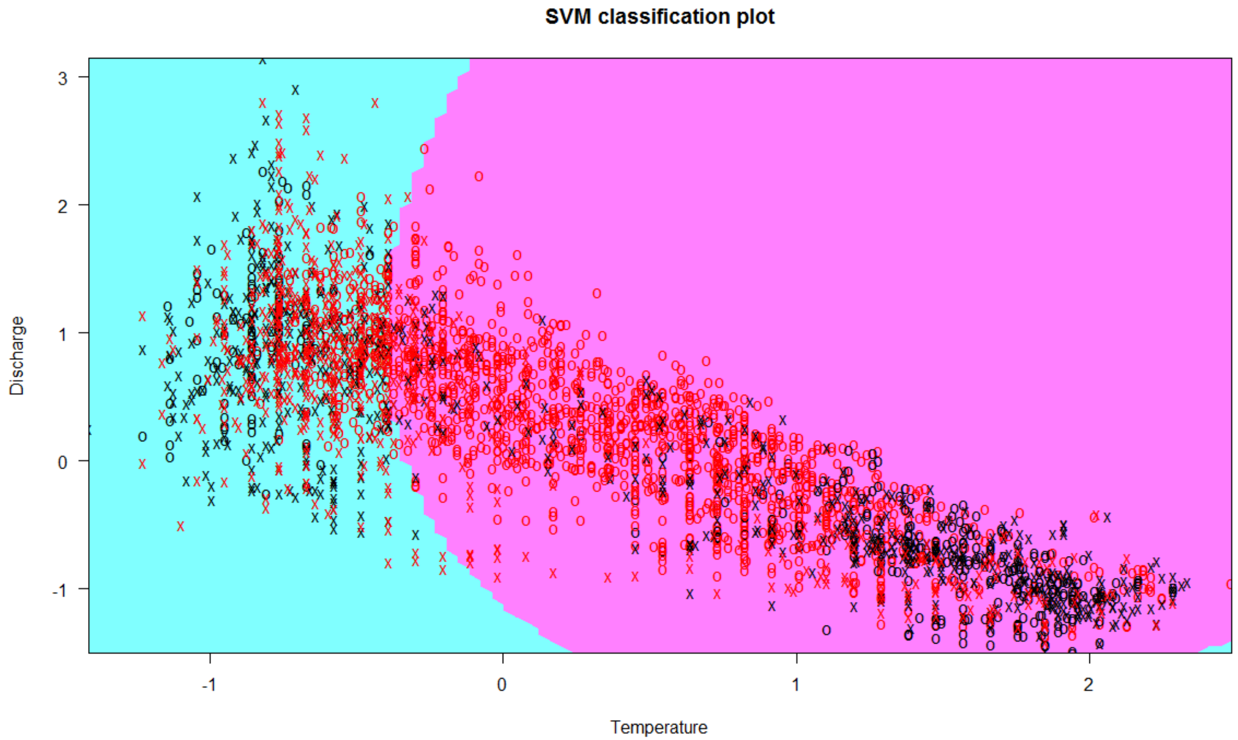
Predicting Running Days

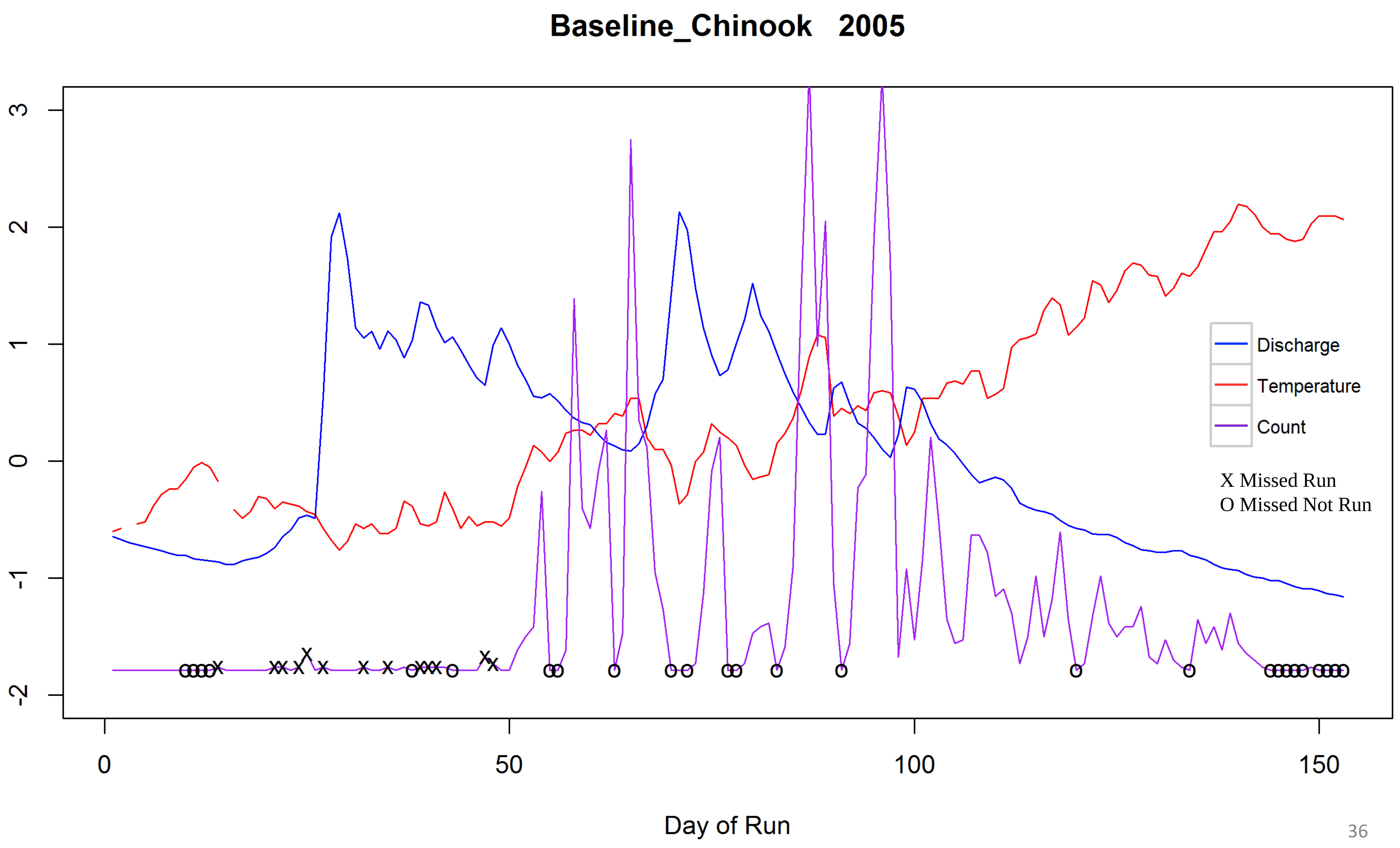
- Use Machine Learning Techniques
  - Learn factors that signal run days
  - Predict the no fish days in the middle of the run
- Variables to try:
  - Temperature and discharge
  - Day of run
  - Change in temperature and discharge
  - History of values

Baseline SVM Performance

Variables: Temperature and discharge

Metric	Average Performance	95% Conf.
Accuracy	83.23%	8.24%
Recall Run	70.51%	29.82%
Recall Not	89.70%	13.94%
Precision	78.88%	24.48%
Precision	86.40%	12.24%





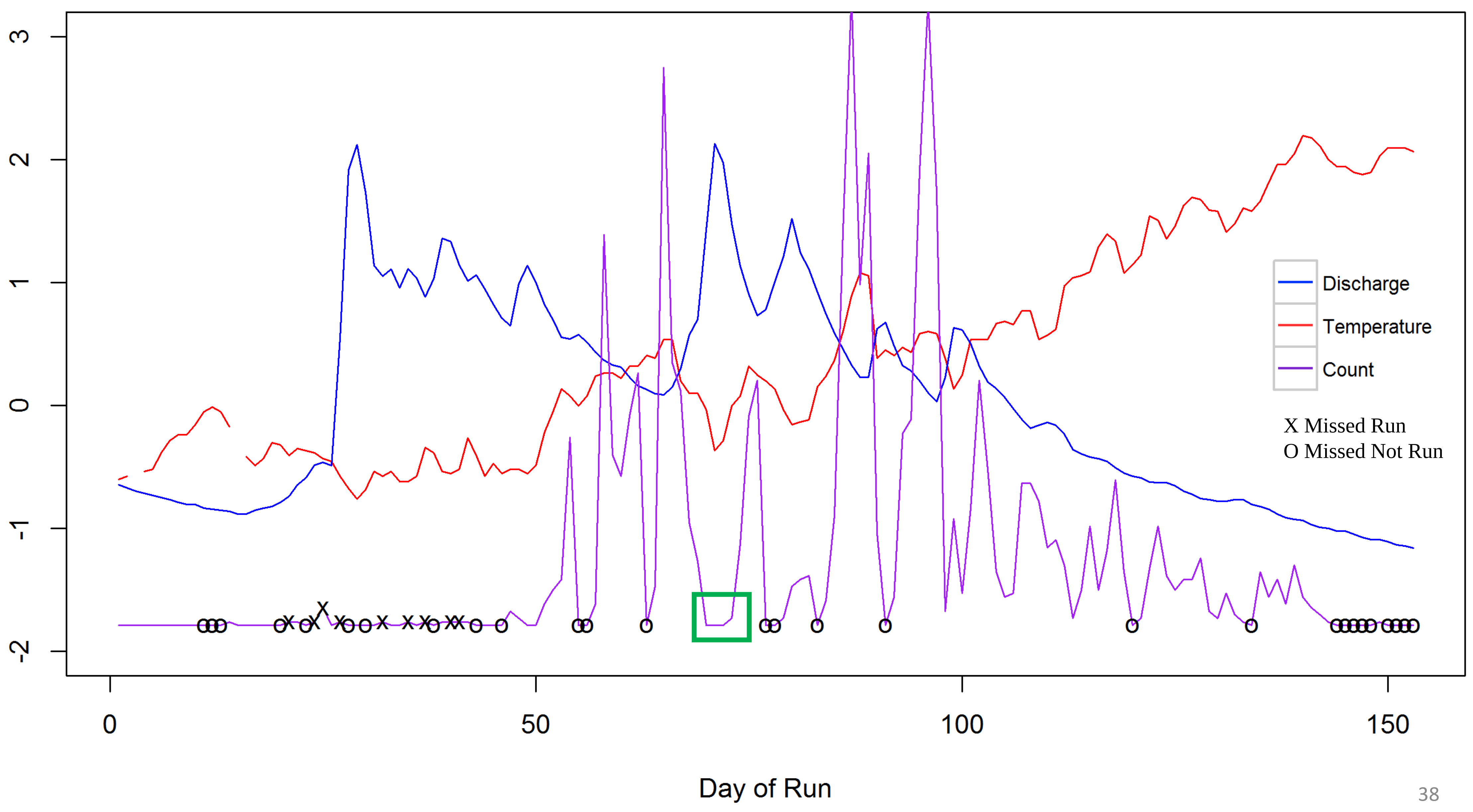
Baseline_Chinook 2005
Discharge
Temperature
Count
X Missed Run
O Missed Not Run
Day of Run
0
50
100
150
-2
-1
0
1
2
3

History SVM Performance

- Variables:
  - Temperature
  - Discharge
  - Derivatives
  - Day of Run
  - 3 Day History

Metric		Average Performance	95% Conf.
Accuracy –	Baseline	83.23%	8.24%
	All Vars	85.26%	7.84%
Recall Run –	Baseline	70.51%	29.82%
	All Vars	69.51%	19.74%
Recall Not –	Baseline	89.70%	13.94%
	All Vars	92.95%	9.12%
Precision Run –	Baseline	78.88%	24.48%
	All Vars	82.34%	21.66%
Precision Not –	Baseline	86.40%	12.24%
	All Vars	86.65%	9.06%

History Chinook 2005



Predicting runs: Conclusions

Predict Median Day of the run within ± 8 days

SVM Failed to learn why 0 days occurs during mid run

Didn’t handle year to year variation in water profile well

Needed to find feature that captures the dynamics that the fish respond to

Future Directions

Probabilistic Model:

- Infer distribution of fish waiting in the Ocean for conditions to be right
- Model distances swam in river

Apply to different species

Differences between wild and hatchery

Thank You

Tom Dietterich

Collaborators: Rebecca Flitcroft and Gordon Grant