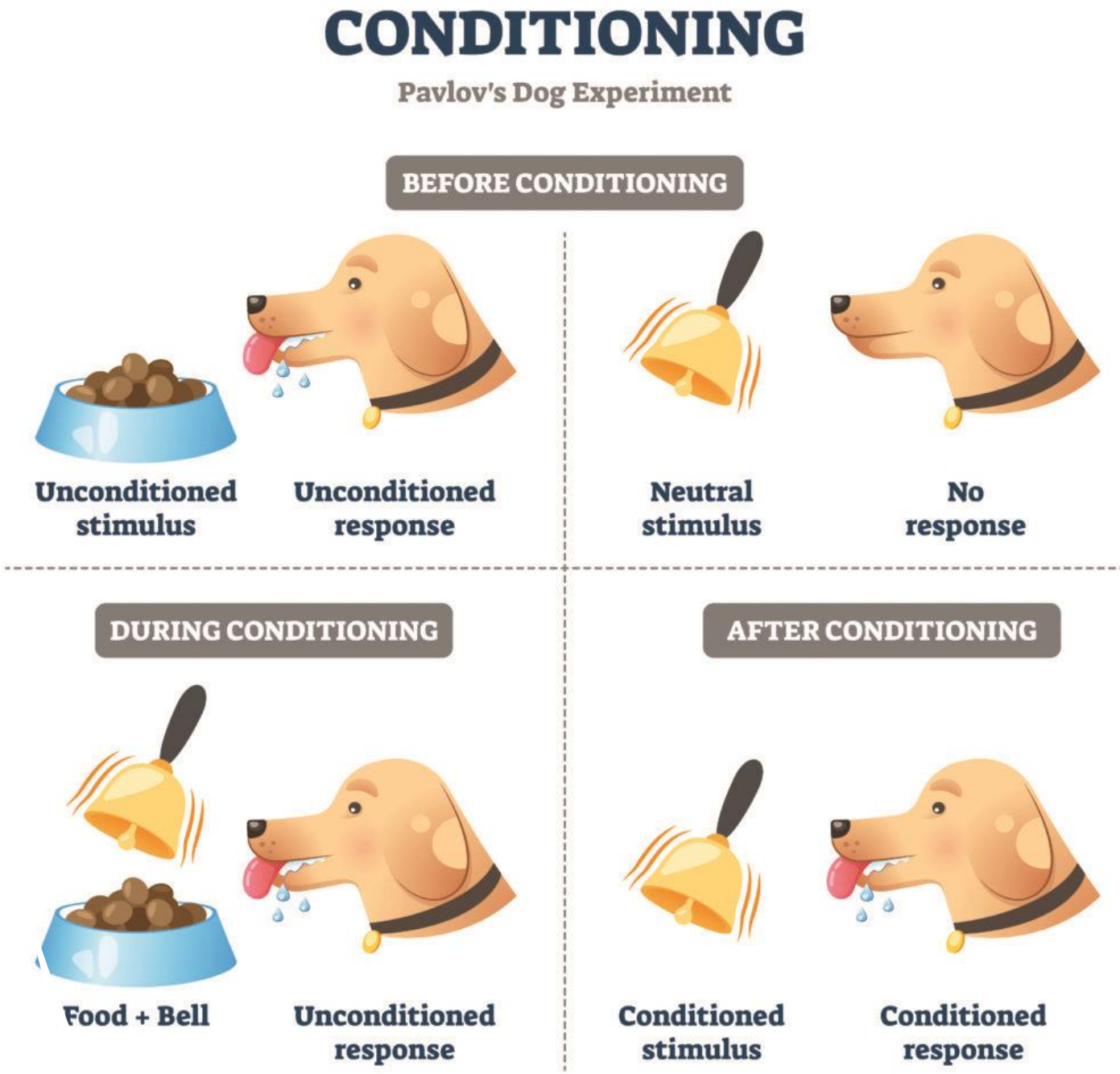
Statistics in Neuroscience

By: David Ye

Mentor: Ethan Ancell

The Experiment

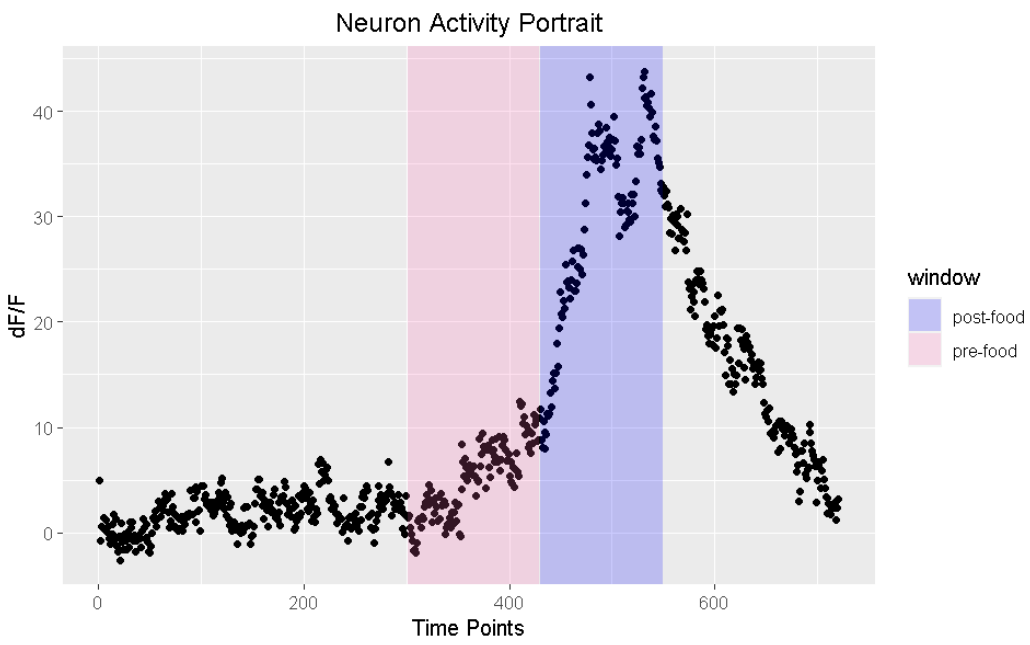
- Classical conditioning on mice
  - Neutral stimulus: two different tones
  - Unconditioned stimulus: food and shock
- Question of Interest
  - Does classical conditioning influence neuron activity?
- 2-Photon Calcium Imaging
  - New tech for looking at calcium levels in neurons (calcium presence signifies neuron firing)
  - The variable dF/F measures calcium presence

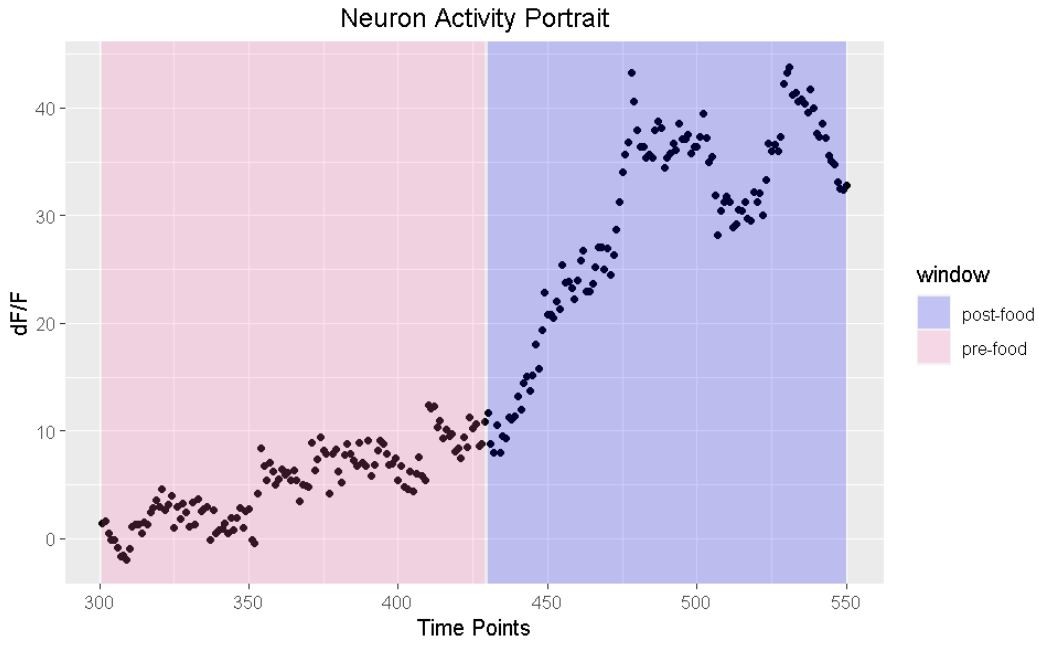


Hypothesis Testing

- In this experiment, we want to see if neuron activity changes after an event occurred (i.e. food is given/tone is played)
- Null Hypothesis: the mean dF/F is the same in the pre-event and post-event windows
- Alternative Hypothesis: the mean dF/F is different in the pre-event and post-event windows

We'll focus on pre/post food windows:



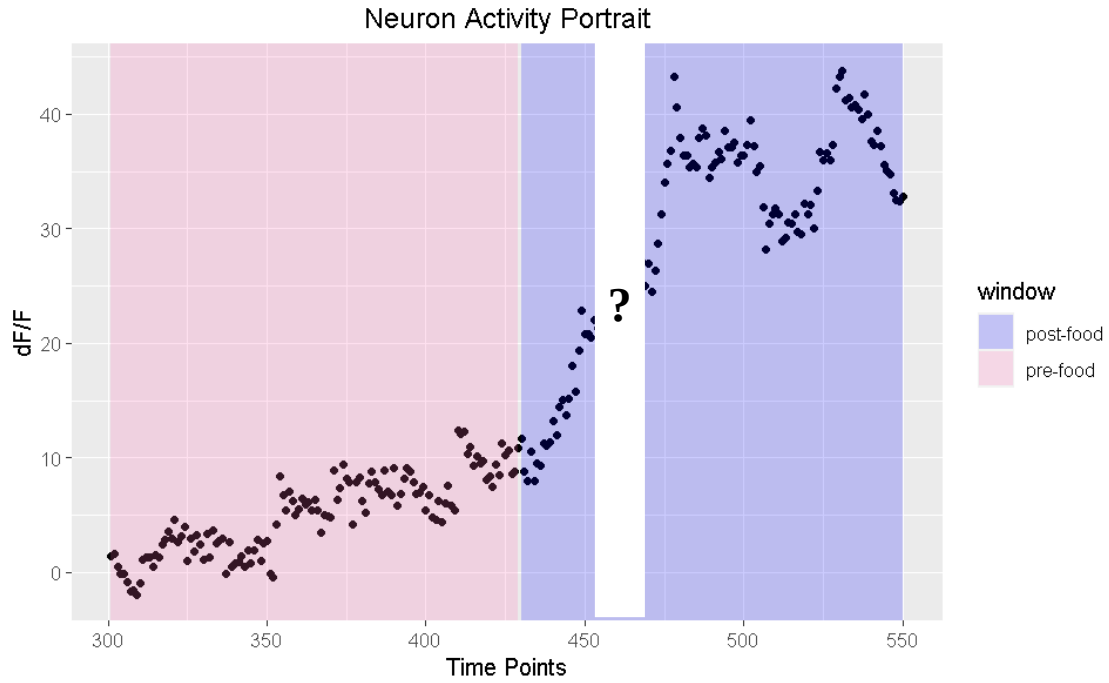


Statistical Analysis:

- Two sample t-test
- Permutation Distribution based Approaches:
  - Permutation Test
  - Uniform Wrapping
  - K-Chunking
  - Criss-Cross with Uniform Wrapping

Naive Approach: Two Sample t-Test

- Well known method for testing differences in means
- Assumption for t-test:
  - Observed dF/F in both pre-food and post-food windows must be obtained via a random sample
- Random Sample
  - A sequence of independent and identically distributed dF/F



$$\textbf{p-value} = 3.358 * 10^{-8}$$

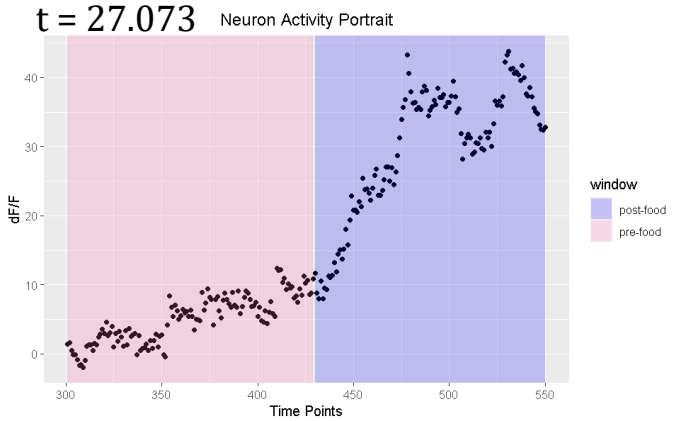
p-value is invalid because assumptions are violated (the t-statistic may follow some unknown distribution)

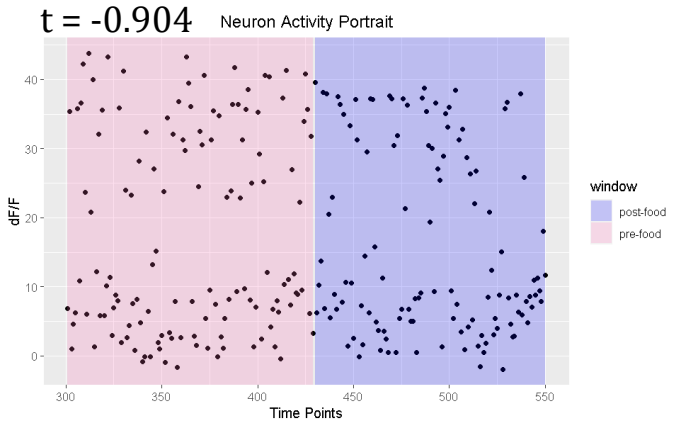
Finding another approach…

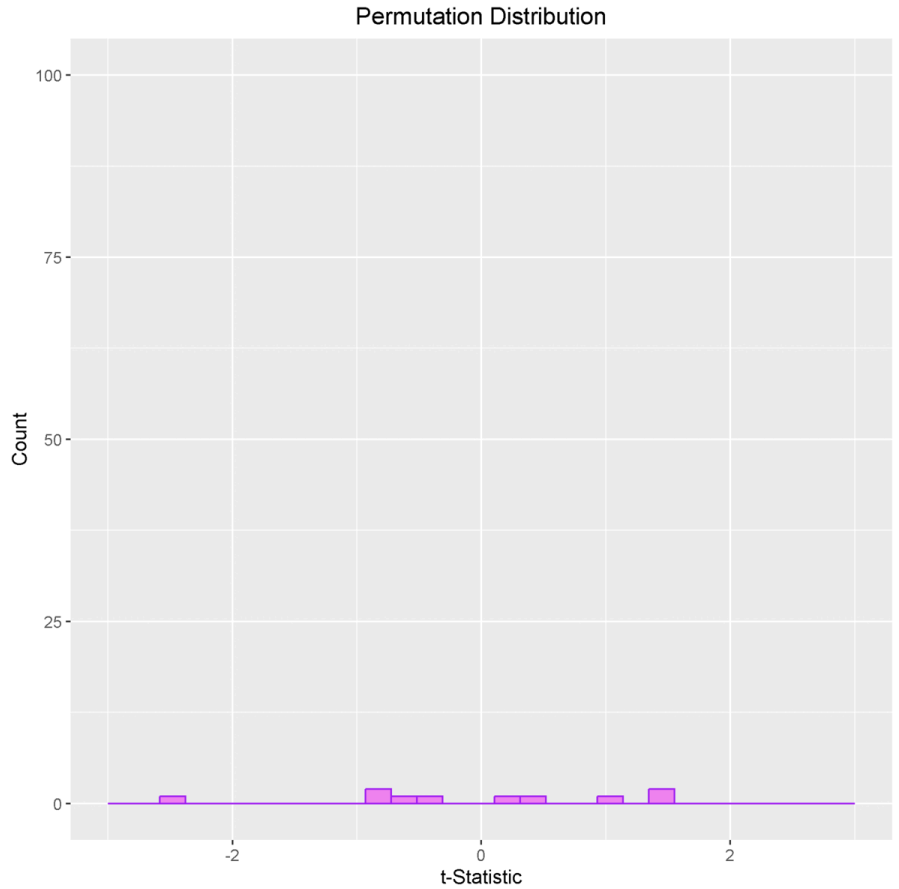
- Can’t use t-distribution so we need another way to obtain p-values
- We will use a **restricted randomization scheme** to approximate the t-statistic under the null hypothesis of no mean neuron difference
- Randomization is a powerful way to approximate unknown distributions
- We will first explore unrestricted randomization and think about ways to restrict randomization

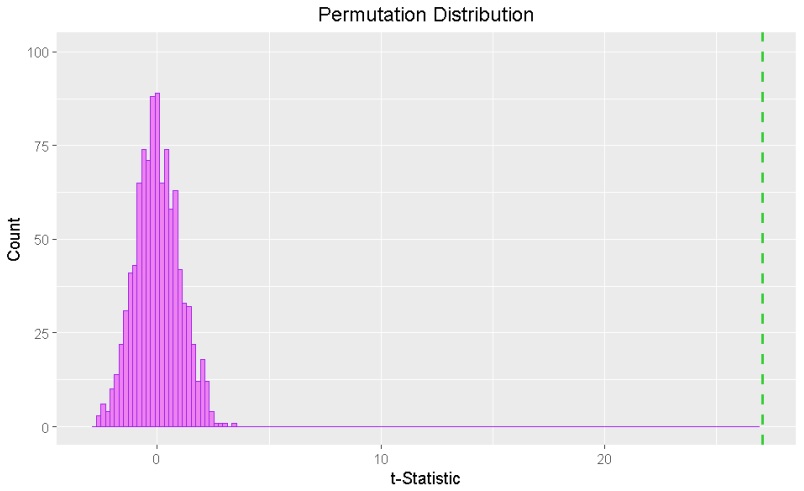
Approach #1: Permutation Test

1. Calculate the t-statistic for the pre-food and post-food window
2. Repeatedly shuffle the dF/F values in pre-food and post-food windows and repeat step 1. at each iteration
3. Obtain p-value by counting the number of t-statistics as or more extreme than our initial t-statistic





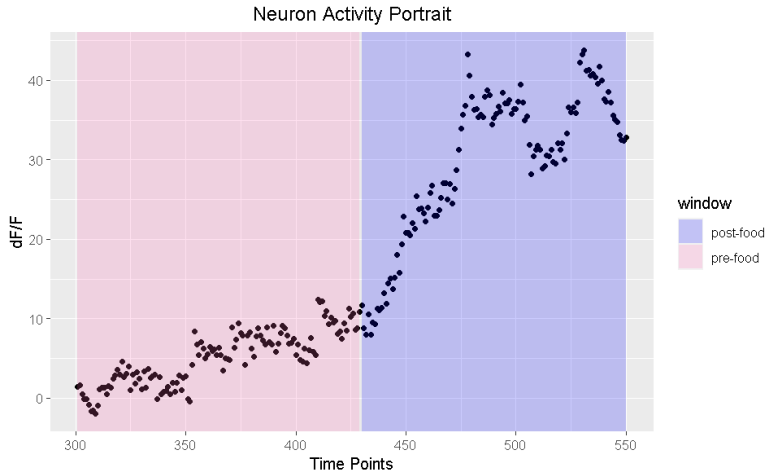


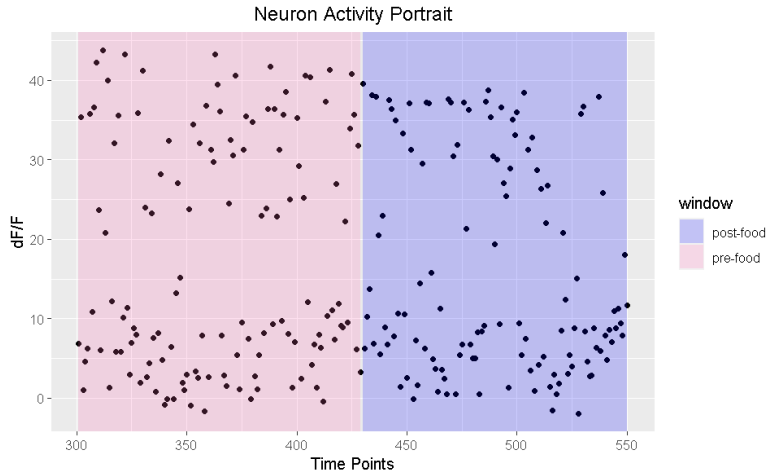


$$\textbf{p-value} = 2\left(\frac{0}{1000}\right) = 0$$

Assumptions for Permutation Tests

- Exchangeability
  - Sequence of dF/F is exchangeable if any permutation of the sequence has the same joint probability distribution
- Check Assumption of Exchangeability
  - Visually:



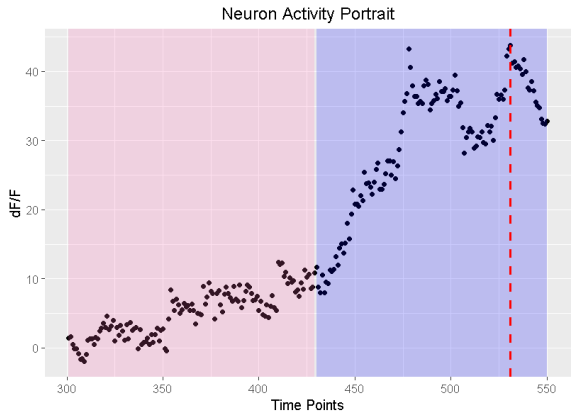


  - $P\left(\frac{dF}{F}_{300}, \frac{dF}{F}_{301}, \frac{dF}{F}_{302}, \cdots, \frac{dF}{F}_{549}, \frac{dF}{F}_{550}\right) \neq P\left(\frac{dF}{F}_{300}, \frac{dF}{F}_{550}, \frac{dF}{F}_{301}, \frac{dF}{F}_{549}, \cdots, \frac{dF}{F}_{425}, \frac{dF}{F}_{426}\right)$
  - Intuition: neurons tend to activate/deactivate gradually rather than jump volatilely between active and inactive states

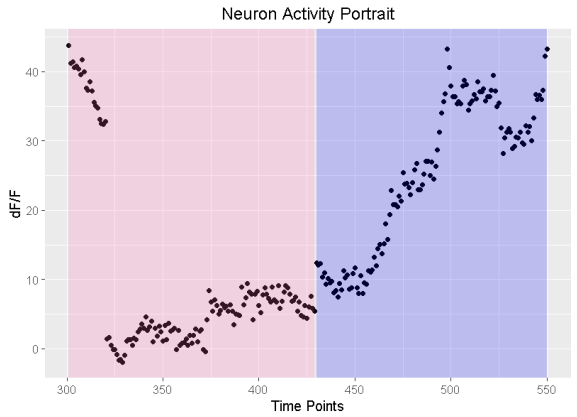
Exchangeability Assumption Violated

Approach #2: Uniform Wrapping

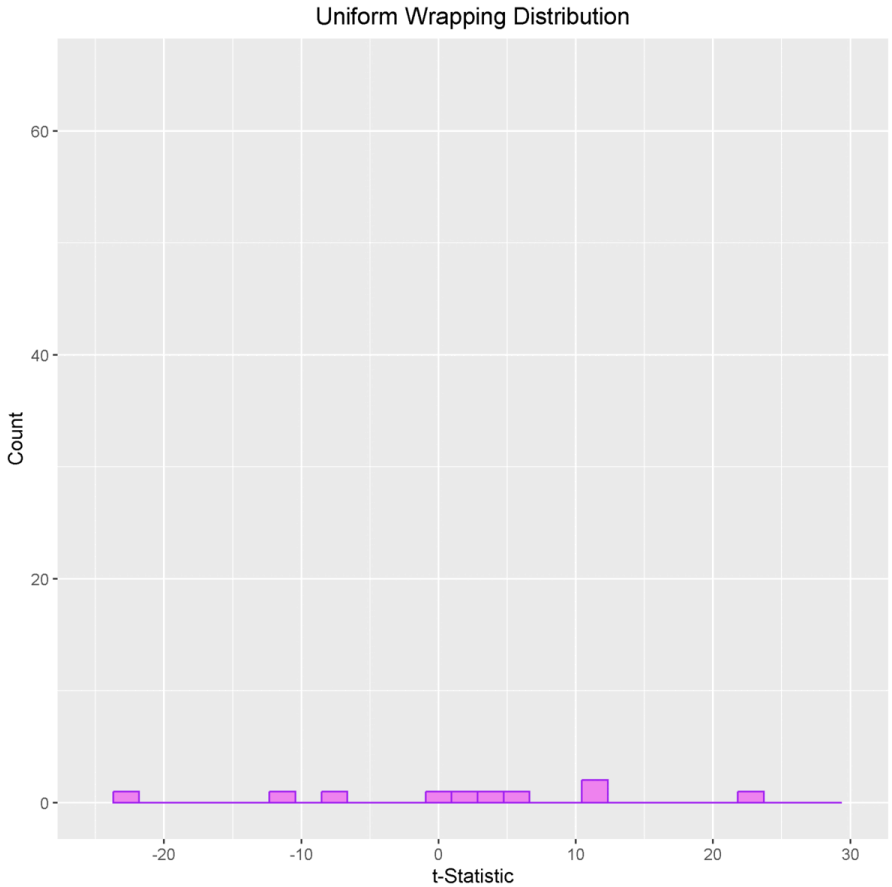
1. Calculate the t-statistic for the pre-food and post-food window
2. Select a uniformly random time point between the pre/post food window
3. Push sequence forward until the selected index is the endpoint and move data pushed out of bounds to the beginning of the sequence and repeat step 1.
4. Repeat steps 2. and 3. on the original data
5. Obtain p-value by counting the number of t-statistics as or more extreme than our initial t-statistic

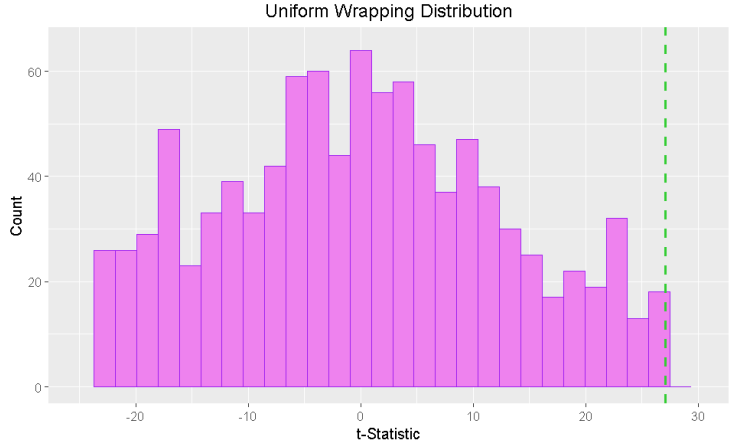


$t = 27.073$



$t = 10.536$





$p\text{-value} = 2(3/1000) = 0.006$

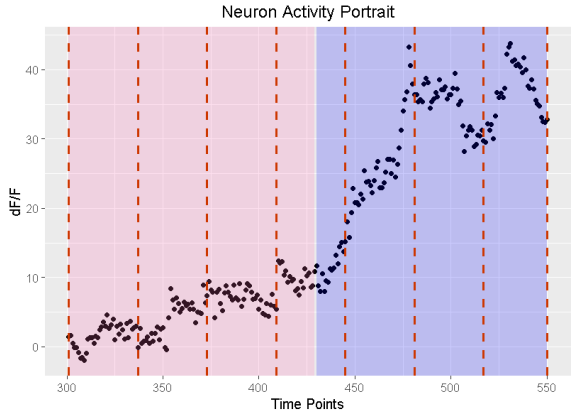
Uniform Wrapping Pros and Cons

- Pros:
  - Random wrapping destroys the association between neural activity and the time of events (this is a way to approximate the null distribution of the t-statistic)
  - Autocorrelation between time points is mostly preserved: this mostly preserves the joint distribution of the data
- Cons:
  - No theoretical basis* (just a clever idea to approximate the null distribution)
  - Limited number of permutations (only 250 possible permutations)
    - Image of data under uniform wrapping does not capture the full spectrum of an unresponsive neuron
  - Overrepresentation of neural traces with peaks if neuron portrait has a peak
- Next approach:
  - K-Chunking: Many more possible permutations of data to draw a null distribution from

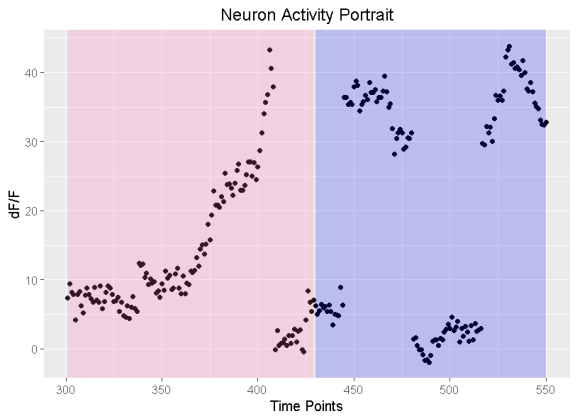
*Uniform Wrapping, K-Chunking, Criss-Cross with Uniform Wrapping all have no theoretical basis

Approach #3: k-Chunking

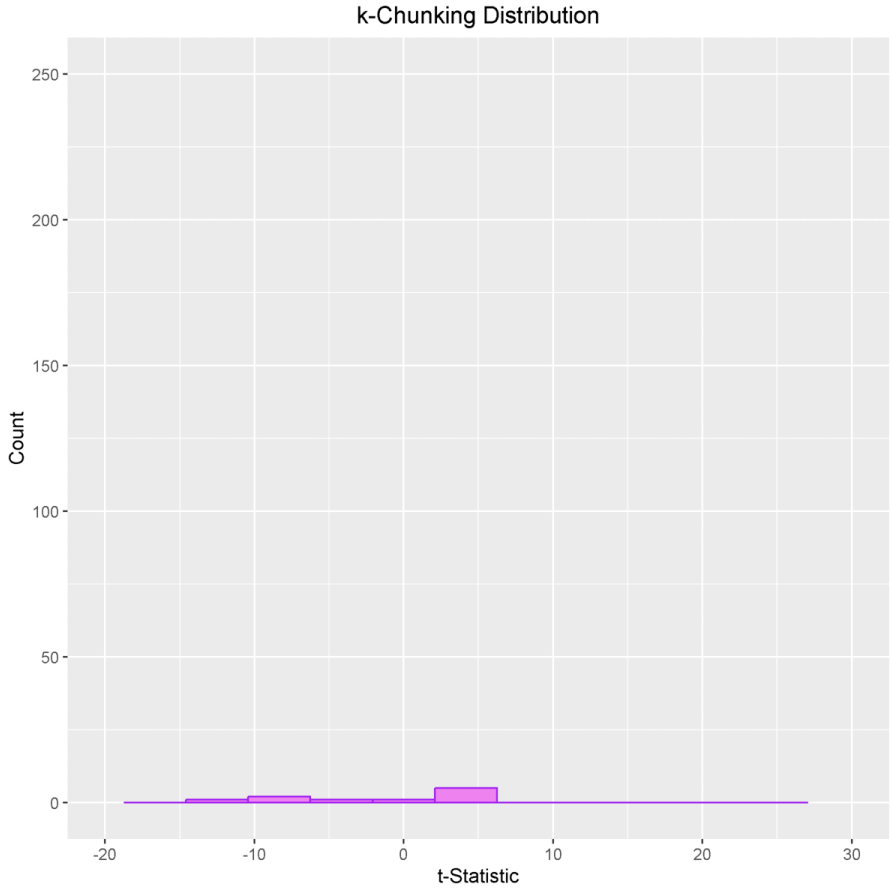
1. Calculate the t-statistic for the pre-food and post-food window
2. Break data into *k* equally spaced intervals
3. Repeatedly shuffle the intervals and repeat step 1. at each iteration
4. Obtain p-value by counting the number of t-statistics as or more extreme than our initial t-statistic

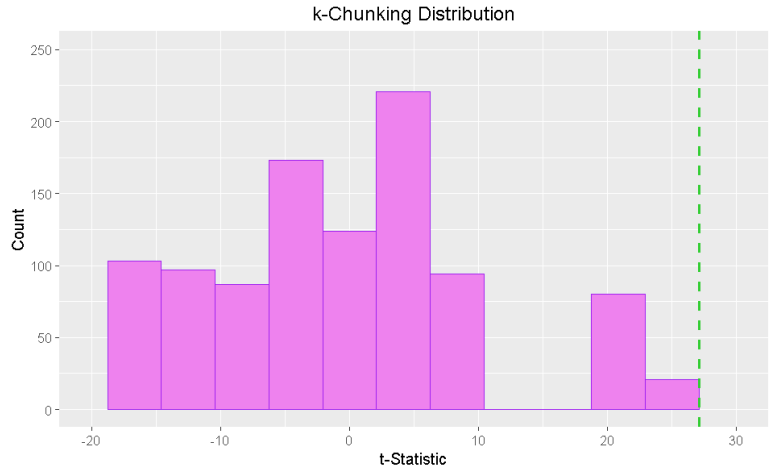


t = 27.073



t = 4.596





p-value = $2(0/1000) = 0$

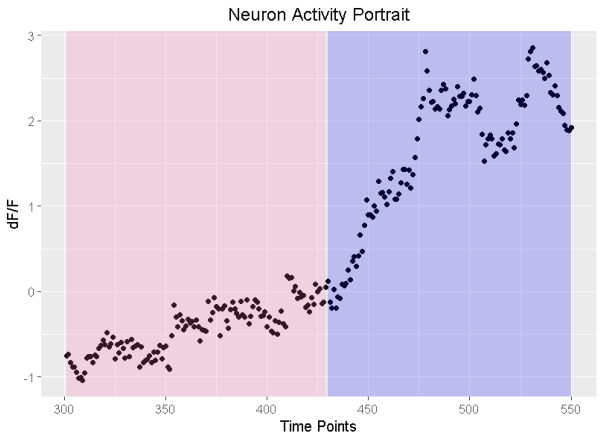
K-Chunking Pros and Cons

- Pros:
  - most of the autocorrelation structure is preserved, only broken up at the edges of the chunks
  - Much larger palette to draw from to approximate the null distribution
- Cons:
  - Overrepresentation of neural traces with peaks if neuron portrait has a peak
  - No theoretical basis* (just a clever idea to approximate the null distribution)
- Next approach:
  - Criss-Cross with Uniform Wrapping: less likely to overrepresent peaks

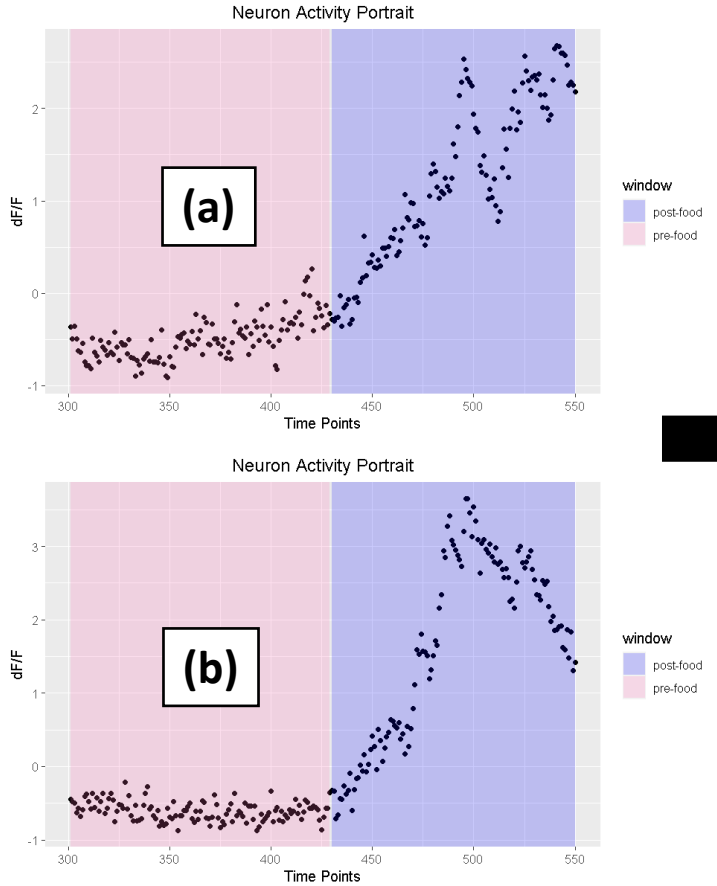
*Uniform Wrapping, K-Chunking, Criss-Cross with Uniform Wrapping all have no theoretical basis

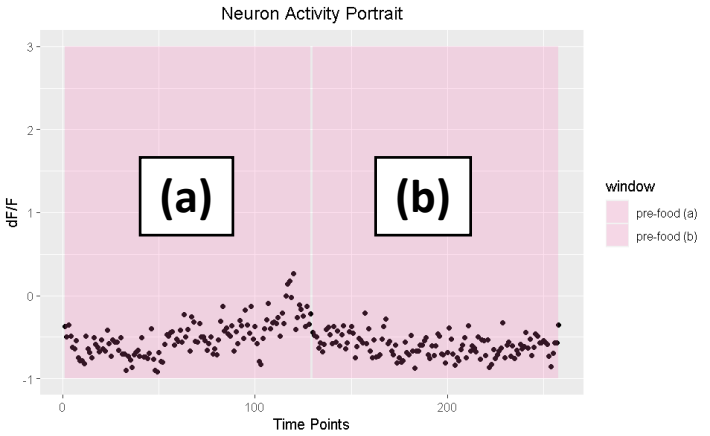
Approach #4: Criss-Cross with Uniform Wrapping

1. Standardize every neuron activity
2. Calculate the t-statistic for the pre-food and post-food window
3. Randomly select two trials of the neuron of interest and combine the pre-food windows from both trials
4. Repeatedly perform uniform wrapping with combined pre-food windows
5. Repeat steps 3. and 4. many times
6. Obtain p-value by counting the number of t-statistics as or more extreme than our initial t-statistic



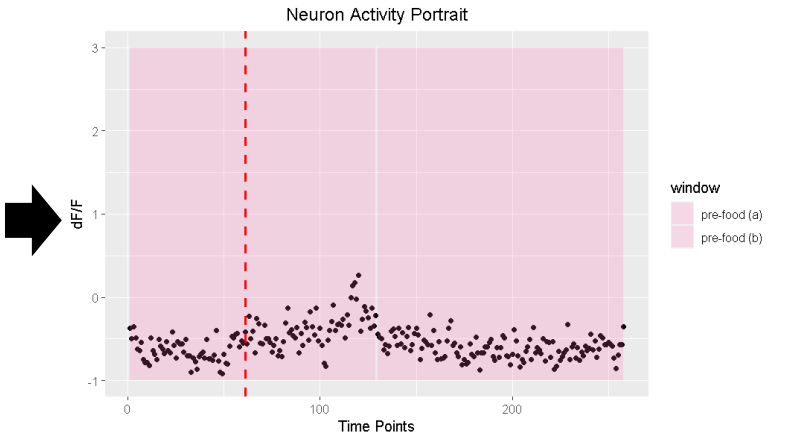
t = 27.073

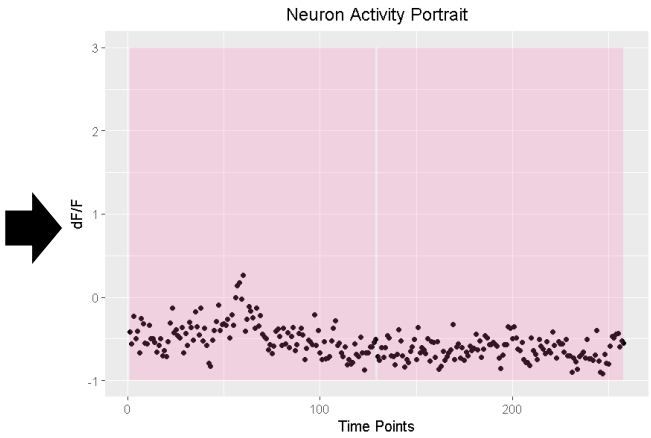




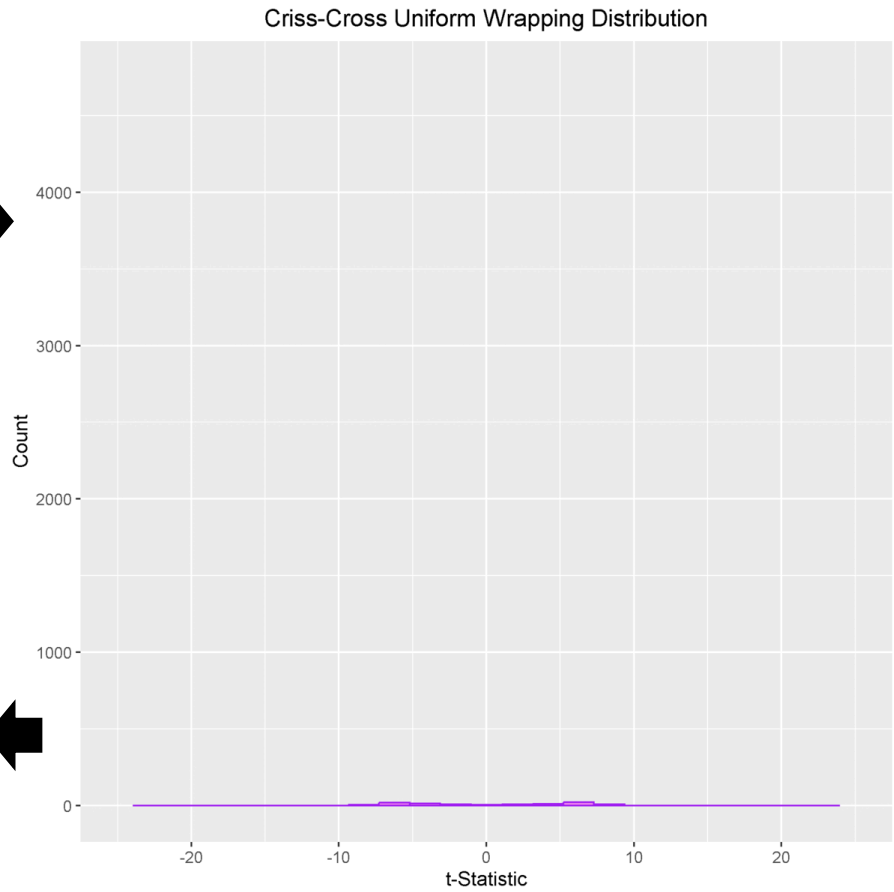
Approach #4: Criss-Cross with Uniform Wrapping

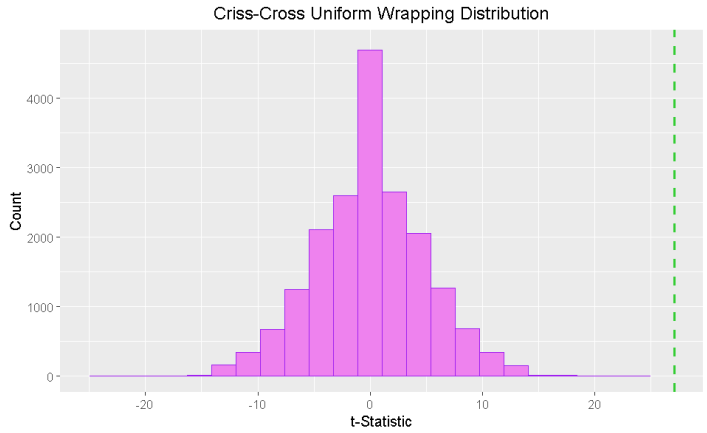
1. Standardize every neuron activity
2. Calculate the t-statistic for the pre-food and post-food window
3. Randomly select two trials of the neuron of interest and combine the pre-food windows from both trials
4. Repeatedly perform uniform wrapping with combined pre-food windows
5. Repeat steps 2. and 3. many times
6. Obtain p-value by counting the number of t-statistics as or more extreme than our initial t-statistic





t = 4.596





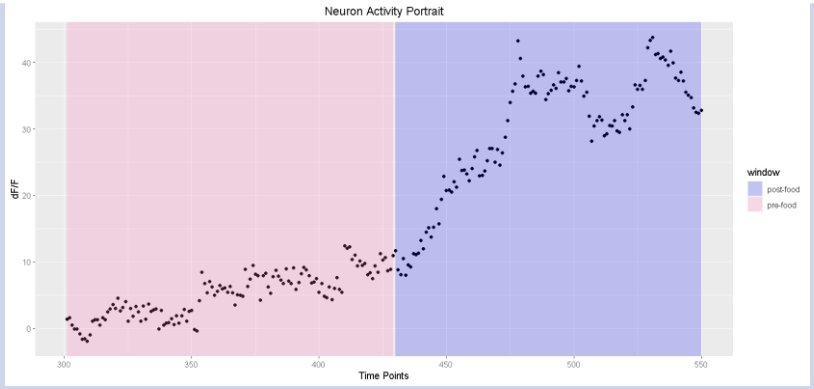
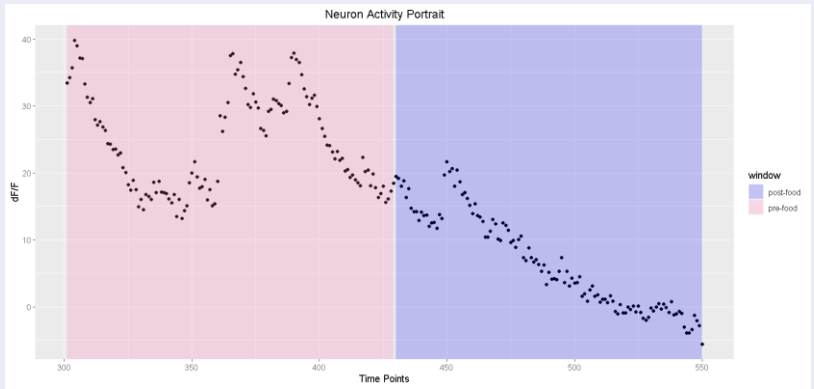
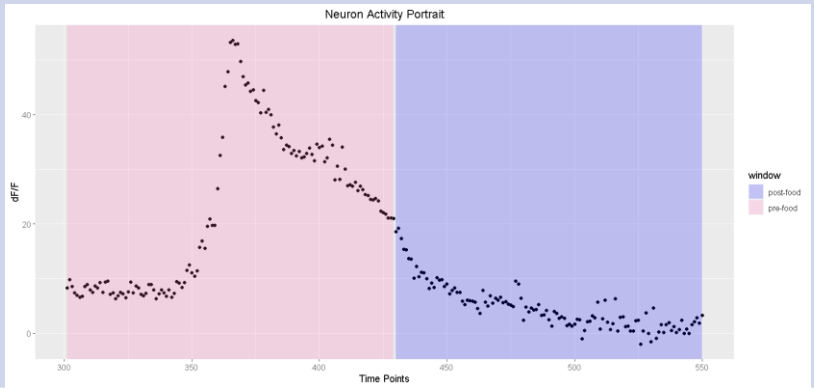
p-value = 2(0/19000) = 0

Criss-Cross with Uniform Wrapping Pros and Cons

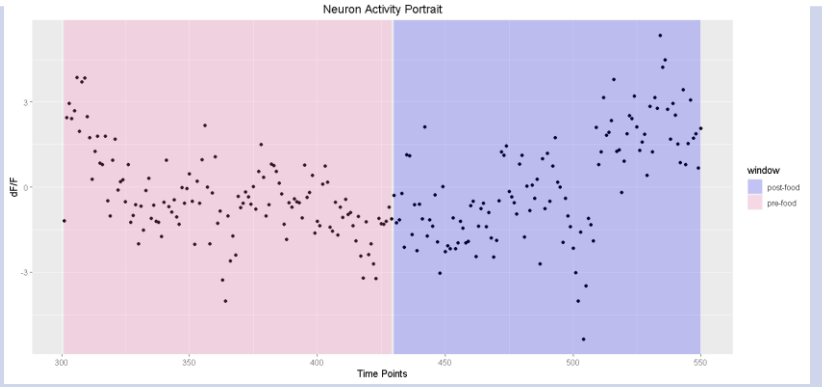
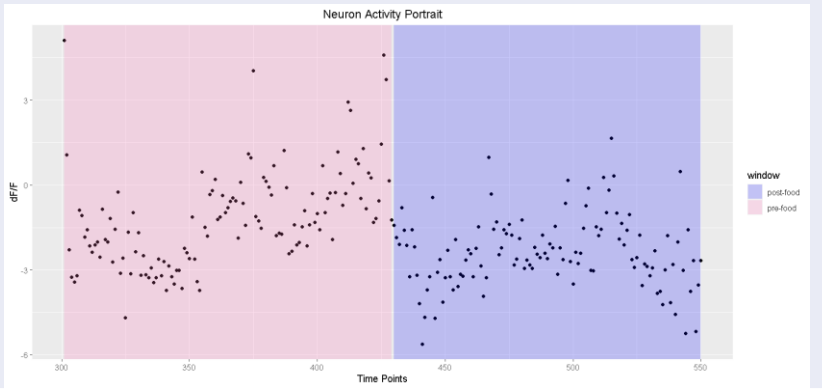
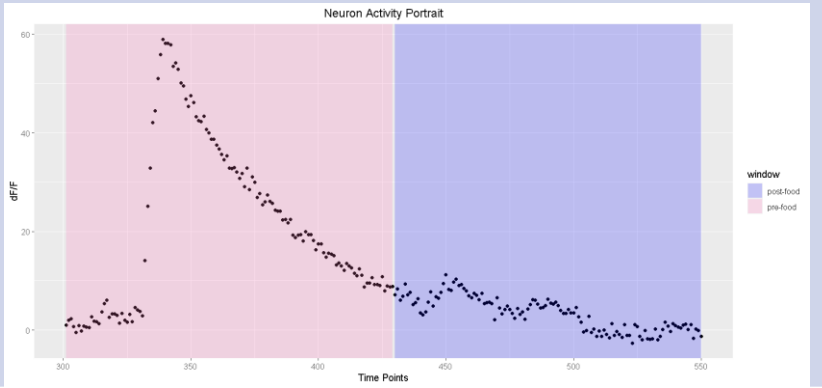
- Pros:
  - autocorrelation between time points is mostly preserved
  - large palette to draw from to approximate the null distribution
  - does not overrepresent neural traces with peaks if neuron portrait has a peak
- Cons:
  - No theoretical basis* (just a clever idea to approximate the null distribution)

*Uniform Wrapping, K-Chunking, Criss-Cross with Uniform Wrapping all have no theoretical basis

P-Value Comparison Table (orange – significant at 95% confidence level)

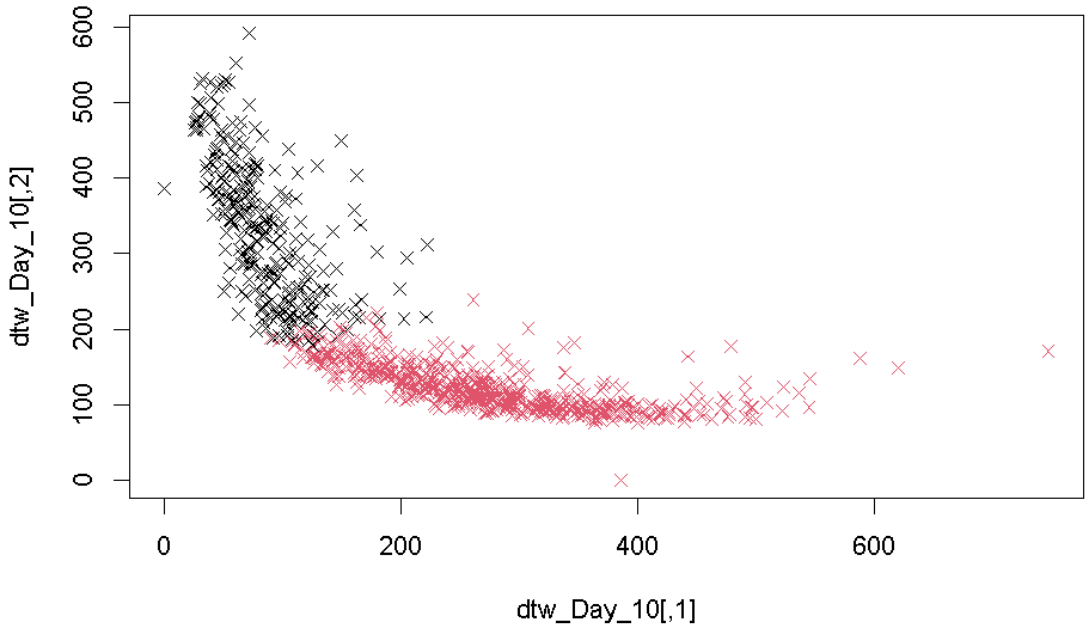
t-Test	Permutation Test	Uniform Wrapping	K-Chunking	Criss-Cross with U.W.	Portrait
$5.438 * 10^{-78}$	0	$4.2 * 10^{-3}$	0	0	
$-2.019 * 10^{-51}$	-0	$-4.8 * 10^{-3}$	$-9.6 * 10^{-2}$	$-2.801 * 10^{-2}$	
$-1.596 * 10^{-31}$	-0	-0.424	$-1.006 * 10^{-1}$	$-1.026 * 10^{-1}$	

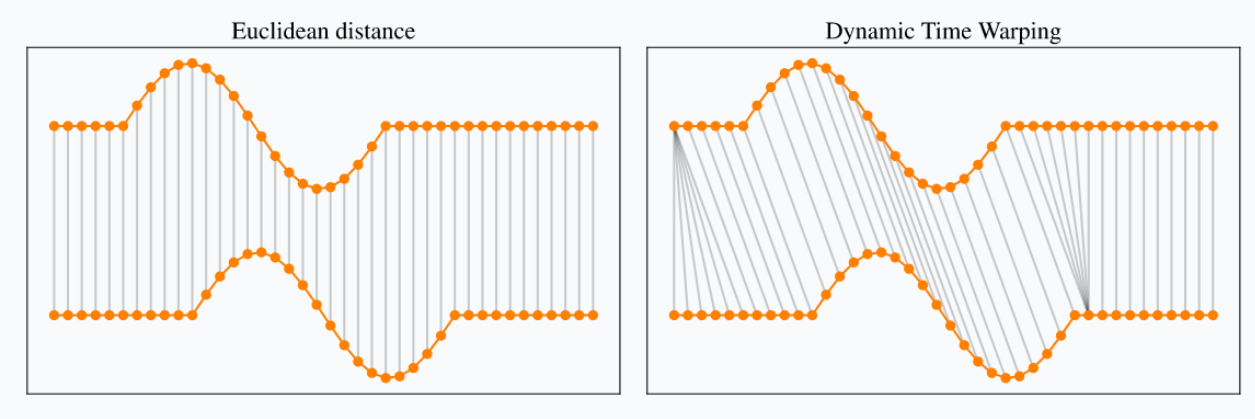
P-Value Comparison Table (orange – significant at 95% confidence level)

t-Test	Permutation Test	Uniform Wrapping	K-Chunking	Criss-Cross with U.W.	Portrait
$3.002 * 10^{-2}$	$3.06 * 10^{-2}$	$8.53 * 10^{-1}$	$5.07 * 10^{-1}$	$7.581 * 10^{-1}$	
$-1.302 * 10^{-9}$	-0	$-4.0 * 10^{-3}$	-0	$-4.214 * 10^{-1}$	
$-2.759 * 10^{-25}$	-0	$-2.994 * 10^{-1}$	$-8.980 * 10^{-2}$	$-3.327 * 10^{-2}$	

Unsupervised ML Approach:

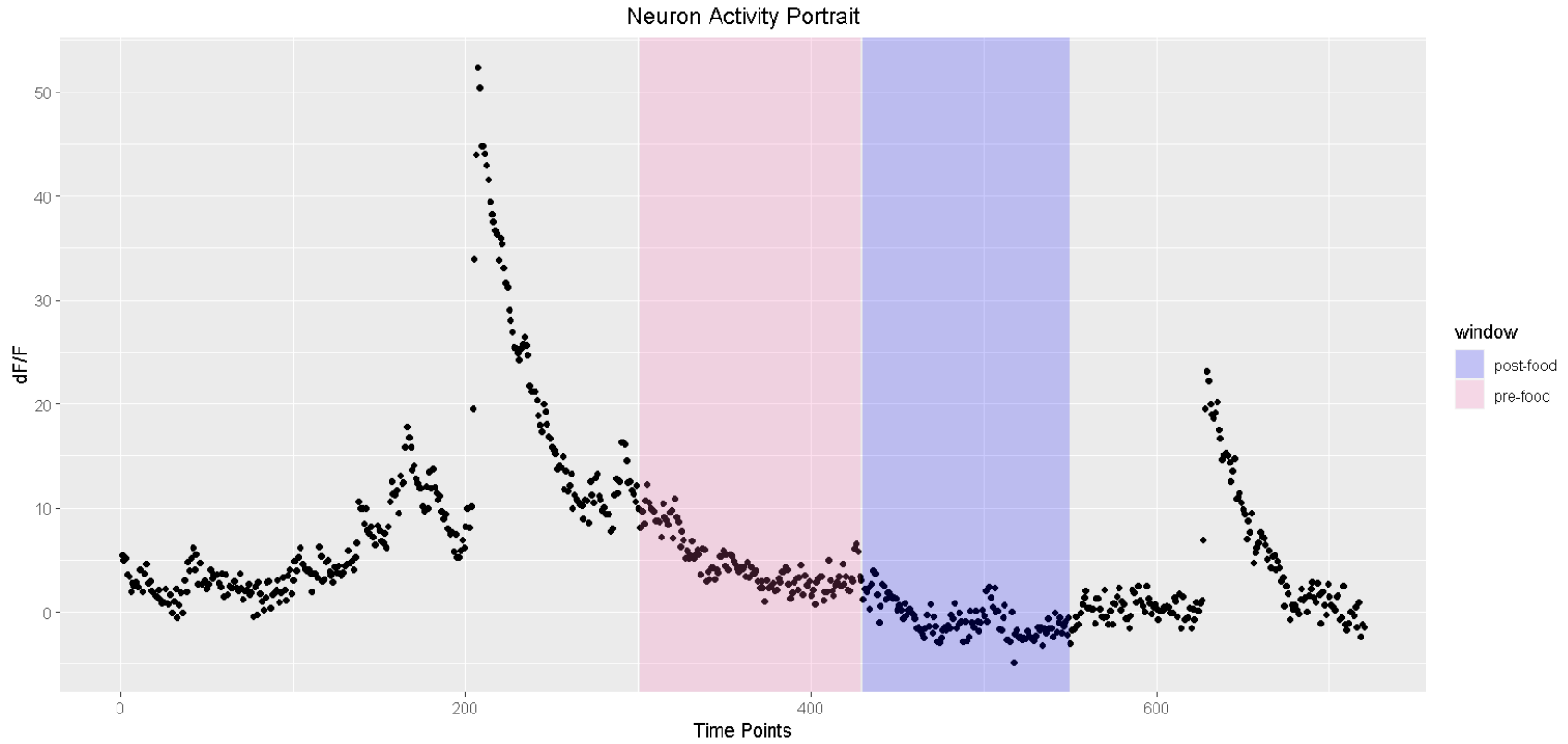
- Spectral Clustering with Dynamic Time Warping
  - A k-means clustering technique based on the Laplacian matrix of a similarity matrix built from the dynamic time warpping distance metric





Discussions

- Is there a time series approach to answering the question “does classical conditioning influence neuron activity”?
- Is there an unsupervised learning approach?
  - Possible approach to try is spectral clustering approach with DTW
- Is there any theoretical basis to the tests being used?
- Is there a better approach than comparing means of pre/post event windows?



Significant p-values for all tests…

Thank You!