# The Cannabinoid Content of Legal Cannabis in Washington State Varies Systematically Across Testing Facilities and Popular Consumer Strains & Products

Replication Data Codebook

## Dataset Description

Replication data for all analyses contained within 'The Cannabinoid Content of Legal Cannabis in Washington State Varies Systematically Across Testing Facilities and Popular Consumer Strains & Products.' Dataset comprises Washington's LCB i502 Recreational Cannabis testing data from June 2014 to May 2017 with additional columns appended from Leafly strain data. Each row is a test result for a recreational cannabis product in the i502 Washington LCB data. Columns with the prefix "strain_" contain values that do not vary within strain and are missing if the test result did not match to a strain in the Leafly.com strain database. The data contain 24 Columns and 304,123 rows.

## Variable Description

### test_id

A unique identifier for each test result contained in the i502 data. Identifiers are chronological.

### test_strain

The strain name as reported by the grower to the lab contained in the i502 data.

### leafly_strain

The strain name in the Leafly.com database of strains that matched to the grower-reported strain name. Values are missing if there was not a close match with a Leafly strain.

### no_match

An indicator variable equal to 1 if the test result did not match to a Leafly.com strain name and equal to 0 if there was a match.

thc_max

Conventionally calculated THC content, as a percentage.

cbd_max

Conventionally calculated CBD content, as a percentage.

date_test

Date and time of the laboratory test as recorded in the i502 data.

year

Convenience column storing the year of the test.

org_name

The grower or manufacturer who submitted the cannabis sample for testing.

org_active

An indicator for whether the organization was active at the time of data retrieval.

lab_name

Name of the laboratory which tested the product.

inventory_type

Text indicating the type of cannabis product.

chemotype

The product's chemotype, as indicated by its THC to CBD ratio. Chemotype 1 is "high-THC," Chemotype 2 is "balanced," and Chemotype 3 is "high-CBD."

strain_category

The Indica, Sativa, or Hybrid category of the test result's strain, as recorded by Leafly.com. Column contains missing values if test did not match to a Leafly strain.

strain_leafly_page_rank

Page view ranking for strain on Leafly.com. A value of 1 indicates the strain with the most page views. Column contains missing values if test did not match to a Leafly strain.

### strain_leafly_review_rank

Review count ranking for strain on Leafly.com. A value of 1 indicates the strain with the most reviews. Column contains missing values if test did not match to a Leafly strain.

### strain_tests

The number of test results containing the strain in the dataset. Column contains missing values if test did not match to a Leafly strain.

### strain_chemotype

The modal chemotype of the strain in the data. Column contains missing values if test did not match to a Leafly strain.

### strain_prop_chemotype1

The proportion of test results for the strain that are Chemotype 1 ("high-THC"). Column contains missing values if test did not match to a Leafly strain.

### strain_prop_chemotype2

The proportion of test results for the strain that are Chemotype 2 ("balanced"). Column contains missing values if test did not match to a Leafly strain.

### strain_prop_chemotype3

The proportion of test results for the strain that are Chemotype 3 ("high-CBD"). Column contains missing values if test did not match to a Leafly strain.

### strain_modal_chemo_prop

The proportion of test results for the strain that fell into the strain's modal chemotype. For example, if the most common category for a strain was Chemotype 1, this column records the proportion of test results for that strain that were Chemotype 1. Column contains missing values if test did not match to a Leafly strain.