

My Projects

My BRCA1 Analysis

My SEER Analysis

1. Browse Data

2. Create Dataset

3. Visualize

4. Analyze

All Cancers > Breast

Source	Subjects	Clinical	Treatment	Demographic	HER2	BRCA
Breast Invasive Carcinoma (TCGA, Nature 2012)	825	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
SEER 1992-2012 San Jose,LA,Rural Georgia and Alaska	14,249	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
SEER 2000-2012 California, Kentucky, Louisiana, New Jersey, Greater Georgia	7,146	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
SEER Louisiana 2005 after Katrina	943	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Confront Cancer Clinic EHR 2014	1458	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

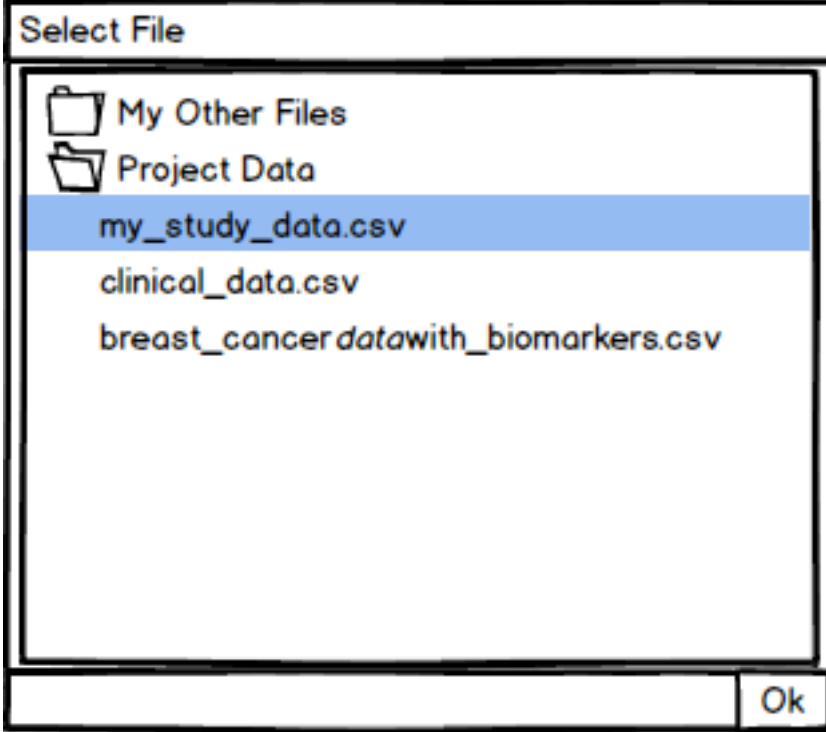
<input type="checkbox"/>	PATIENT ID	AJCC Stage	Converted Stage	Diagnosis	HER2 Final Status	Metastasis-C	miRNA Clu	Mutation C	Overall Survival	Overall Survival (months)	PAM50 Subtype	Person Gender
<input type="checkbox"/>	TCGA-A1-A-001	Stage IIA	Stage IIA	59	Negative	Negative	6	32	LIVING	14.4	Luminal A	FEMALE
<input checked="" type="checkbox"/>	TCGA-A1-A-002	Stage I	Stage I	56	Negative	Negative	6	14	LIVING	43.4	Luminal A	FEMALE
<input checked="" type="checkbox"/>	TCGA-A1-A-003	Stage IIA	Stage IIA	54	Negative	Negative	6		LIVING	48.1		FEMALE
<input checked="" type="checkbox"/>	TCGA-A1-A-004	Stage IIB	Stage IIB	61	Negative	Negative	5		LIVING	14.2		FEMALE
<input type="checkbox"/>	TCGA-A1-A-005	Stage IIA	Stage IIA	39	Negative	Negative	4	78	LIVING	47.2	Luminal A	FEMALE
<input type="checkbox"/>	TCGA-A1-A-006	Stage IIB	Stage IIB	52	Negative	Negative	7		LIVING	20.8		FEMALE
<input checked="" type="checkbox"/>	TCGA-A1-A-007	Stage IIIA	Stage IIIA	39	Negative	Negative	4	29	LIVING	14	Luminal A	FEMALE
<input type="checkbox"/>	TCGA-A1-A-008	Stage IIA	Stage IIA	54	Negative	Negative	5	50	DECEASED	31.8	Basal-like	FEMALE

Upload data from .csv

Upload

Combine data from different data sets into a single data set for analysis.

Add to Custom Dataset



Here we combine data, hone columns to just what we want to look at, merge columns if we have selections from different data sets.

My Projects

My BRCA1 Analysis

My SEER Analysis

My Datasets

My Dataset1

Upload

1 Browse Data2 Create Dataset3 Visualize4 Analyze

PATIENT ID	AJCC Stage	Converted Stage	Diagnosis Age	HER2 Final Status	Metastasis-Coded	miRNA Clusters	Mutation Count	Overall Survival Score	Overall Survival (Months)	PAM50 Subtype	Person Gender
TCGA-A1-A0	Stage I	Stage I	56	Negative	Negative	6	14	LIVING	43.4	Luminal A	FEMALE
TCGA-A1-A0	Stage IIA	Stage IIA	54	Negative	Negative	6		LIVING	48.1		FEMALE
TCGA-A1-A0	Stage IIB	Stage IIB	61	Negative	Negative	5		LIVING	14.2		FEMALE
TCGA-A1-A0	Stage IIIA	Stage IIIA	39	Negative	Negative	4	29	LIVING	14	Luminal A	FEMALE
TCGA-A1-A0	Stage IIA	Stage IIA	54	Negative	Negative	5	50	DECEASED	31.8	Basal-like	FEMALE
TCGA-A1-A0	Stage IIA	Stage IIA	77	Positive	Negative	4	24	LIVING	7.95	Luminal B	MALE

My Dataset1

325 observations of 12 variables

Name	Type
PATIENT ID	identifier
AJCC Stage	categorical
Converted Stage	categorical
Diagnosis Age	numeric
HER2 Final Status	categorical
Metastasis-Coded	categorical
miRNA Clusters	numeric

Diagnosis Age: numeric

Range: 21 - 77

Mean: 44

Median: 54

Std. Deviation: 13.56

Merge With...

miRNA Clusters

Mutation Count

Overall Survival

Merge

Removes this variable from the dataset. Click.

Remove

Histogram

Variable Distribution

Count of Diagnosis Age for each Diagnosis Age (n=1)

Diagnosis Age Bin	Count
20-25	1
25-30	1
30-35	1
35-40	1
40-45	11
45-50	21
50-55	40
55-60	28
60-65	17
65-70	19
70-75	11
75-80	13
80-85	4
85-90	2

My Projects

My BRCA1 Analysis

1 Browse Data

2 Create Dataset

3 Visualize

4 Analyze

Here we combine data, hone columns to just what we want to look at, merge columns if we have selections from different data sets, and save to "My Datasets".

My Datasets

My Dataset1

Upload

PATIENT ID	AJCC Stc	Converted St	Diagnosis /	HER2 Final St	Metastasis-Co	miRNA Clus	Mutation Co	Overall Survival S	Overall Survival (Mc	PAM50 Subt	Person Gen
TCGA-A1-A0	Stage I	Stage I	56	Negative	Negative	6	14	LIVING	43.4	Luminal A	FEMALE
TCGA-A1-A0	Stage IIA	Stage IIA	54	Negative	Negative	6		LIVING	48.1		FEMALE
TCGA-A1-A0	Stage IIB	Stage IIB	61	Negative	Negative	5		LIVING	14.2		FEMALE
TCGA-A1-A0	Stage IIIA	Stage IIIA	39	Negative	Negative	4	29	LIVING	14	Luminal A	FEMALE
TCGA-A1-A0	Stage IIA	Stage IIA	54	Negative	Negative	5	50	DECEASED	31.8	Basal-like	FEMALE
TCGA-A1-A0	Stage IIA	Stage IIA	77	Positive	Negative	4	24	LIVING	7.95	Luminal B	MALE

My Dataset1

325 observations of 12 variables

Name	Type	Remove	Impute Missing Values
PATIENT ID	identifier	<input type="checkbox"/>	<input type="checkbox"/>
AJCC Stage	categorical	<input type="checkbox"/>	<input type="checkbox"/>
Converted Stage	categorical	<input type="checkbox"/>	<input type="checkbox"/>
HER2 Final Status	categorical	<input type="checkbox"/>	<input type="checkbox"/>
Metastasis-Coded	categorical	<input type="checkbox"/>	<input type="checkbox"/>
miRNA Clusters	numeric	<input type="checkbox"/>	<input type="checkbox"/>

HER2 Final Status

Categories:

Negative

Positive

Merge With...

AJCC Stage

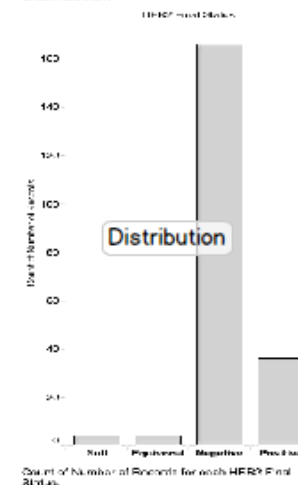
Converted Stage

Metastasis-Cod

Merge

Remove

Distribution



My Projects
My BRCA1 Analysis
My SEER Analysis

Here we select variables and look at box plots, histograms, scatterplots, and covariance matrices.

My Datasets
My Dataset1

Upload Data

1 Browse Data 2 Create Dataset 3 Visualize 4 Analyze

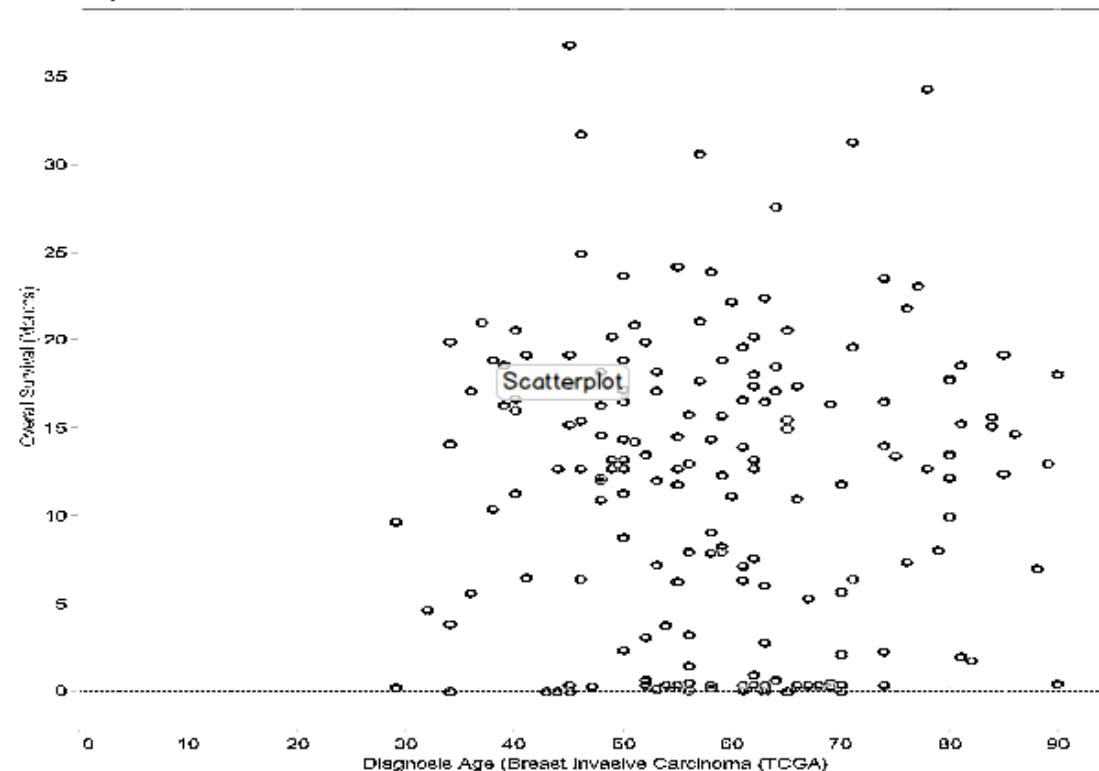
My Dataset1
325 observations of 12 variables

Name	Type
PATIENT ID	identifier
AJCC Stage	categorical
Converted Stage	categorical
Diagnosis Age	numeric
HER2 Final Status	categorical
Metastasis-Coded	categorical
miRNA Clusters	numeric
Mutation Count	numeric
Overall Survival Status	categorical
Overall Survival (Months)	numeric
PAM50 Subtype	categorical
Person Gender	categorical

2 variable scatterplot
2 variable box plot

Plot

Scatterplot



My Projects
My BRCA1 Analysis
My SEER Analysis

Here we select variables and look at box plots, histograms, scatterplots, and covariance matrices.

My Datasets
My Dataset1

Upload Data

1 Browse Data 2 Create Dataset 3 Visualize 4 Analyze

My Dataset1

325 observations of 12 variables

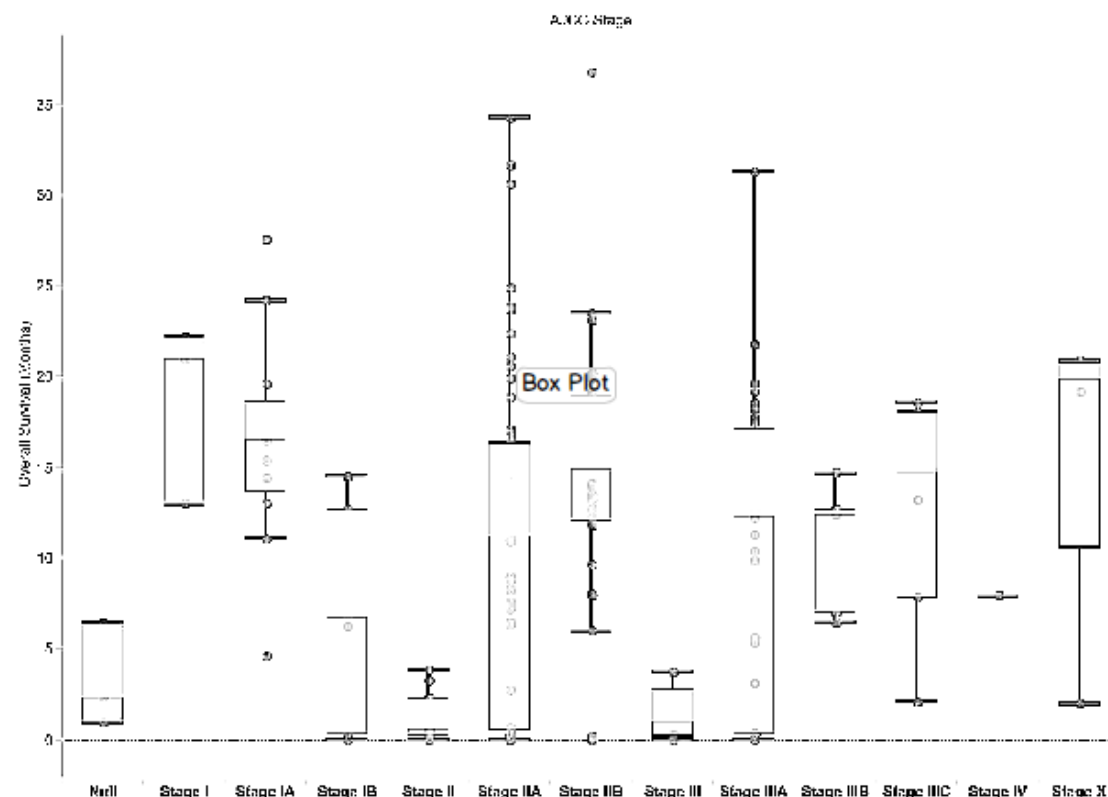
Name	Type
PATIENT ID	identifier
AJCC Stage	categorical
Converted Stage	categorical
Diagnosis Age	numeric
HER2 Final Status	categorical
Metastasis-Coded	categorical
miRNA Clusters	numeric
Mutation Count	numeric
Overall Survival Status	categorical
Overall Survival (Months)	numeric
PAM50 Subtype	categorical
Person Gender	categorical

2 variable box plot ▼

Plot

Select 2 variables from grid on left

BoxPlot



Here we identify the dependent variable, select the type of analysis, do variable transformation, select predictors (if required), view output models including visualizations if available, and save our results.

My Datasets
My Dataset1

Upload Data

1 Browse Data 2 Create Dataset 3 Visualize 4 Analyze

Survival Modeling

Impute Missing Values



* Select Status Variable

* Select Time Variable

Group By:

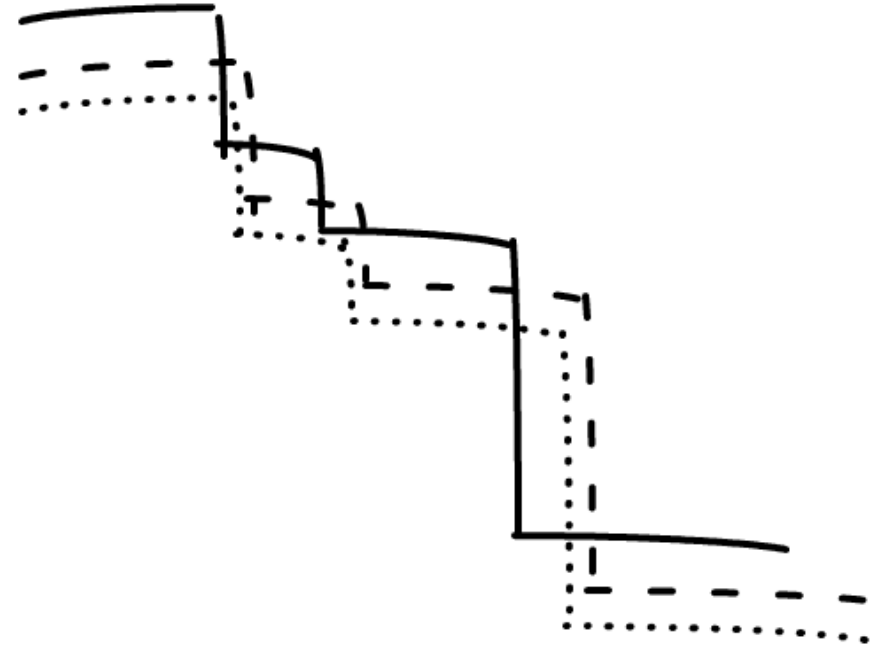
Calculate

HER Status

Survival Month	HER2+/HF	HER2+/HI	HER2-/HF	Triple Negati	Unknow
0	0.995	0.995	0.998	0.996	0.993
1	0.993	0.990	0.997	0.991	0.989
2	0.991	0.986	0.995	0.987	0.986
3	0.990	0.983	0.994	0.982	0.984
4	0.989	0.978	0.993	0.980	0.983
5	0.988	0.975	0.992	0.974	0.982
6	0.987	0.974	0.991	0.969	0.981
7	0.986	0.970	0.990	0.967	0.980
8	0.985	0.969	0.989	0.962	0.979
9	0.985	0.965	0.988	0.958	0.978
10	0.984	0.963	0.987	0.951	0.977
11	0.983	0.962	0.986	0.947	0.976
12	0.982	0.961	0.986	0.943	0.975

Percent Surviving

Date



Save to Project

Here we identify the dependent variable, select the type of analysis, do variable transformation, select predictors (if required), view output models including visualizations if available, and save our results.

My Datasets
My Dataset1

Upload Data

1 Browse Data 2 Create Dataset 3 Visualize 4 Analyze

Survival Modeling

Impute Missing Values



Logistic Regression
Decision Tree
Regression

Select Dependent Variable...
Overall Survival Status

Predictors

☐ Include Second Order Terms

Name	Type
PATIENT ID	identifier
AJCC Stage	
Converted Stage	categorical
Diagnosis Age	numeric
HER2 Final Status	categorical
Metastasis-Coded	categorical
miRNA Clusters	numeric
Mutation Count	numeric
Overall Survival Status	categorical
Overall Survival (Months)	numeric
PAM50 Subtype	categorical
Person Gender	categorical

Multi-select predictor variables

Run

Results

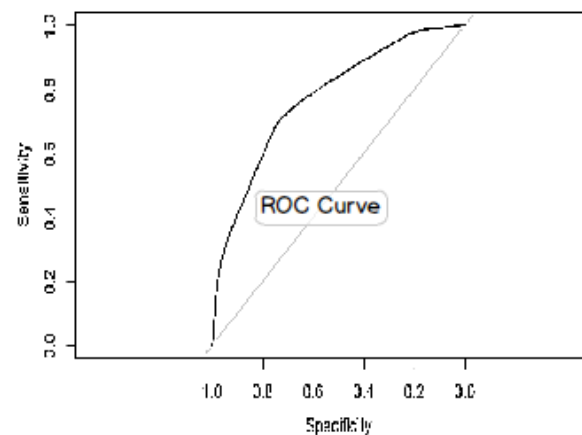
Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.4100689	0.0284684	119.785	< 2e-16 ***
AGE_DX	0.0059060	0.0002772	21.303	< 2e-16 ***

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 633372 on 620335 degrees of freedom
Residual deviance: 521819 on 620320 degrees of freedom
AIC: 521851

Number of Fisher Scoring iterations: 6



Save to Project