The dataset used in this study was acquired from the collection and metadata downloads on the PROVe-AI page from The International Skin Imaging Collaboration. Downloading the metadata file generates a csv with the form: 'prove-ai_metadata_currentyear_currentmonth_currentday.csv'. The collection option downloads a zipped folder with the image data needed for this project with the following name: 'ISIC-images.zip'.

The data dictionary for 'prove-ai_metadata_currentyear_currentmonth_currentday.csv' is as follows:

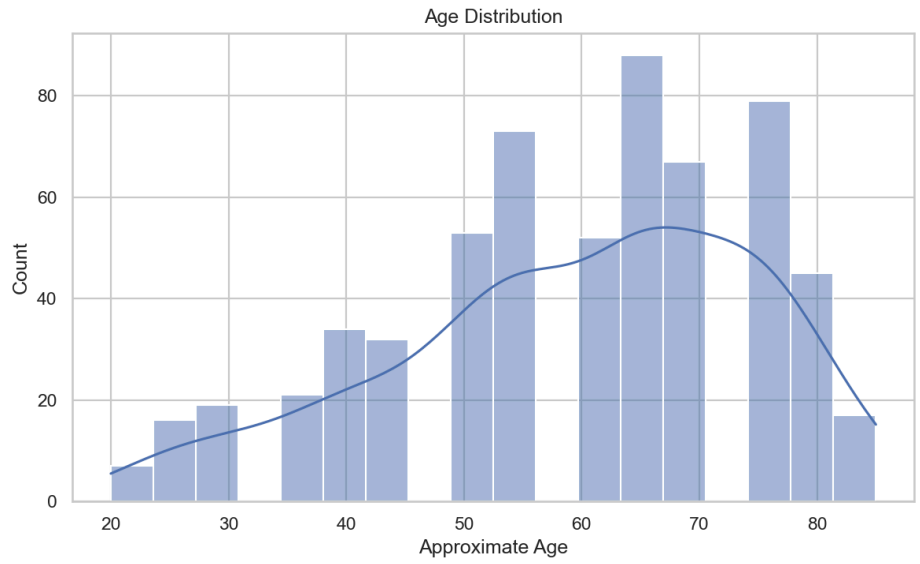
Field Name	**Information Obtained**	**Data Type**
isic_id	Primary key: Unique id for image	string
filename	Name of uploaded file	string
diagnosis_1	Super Category for diagnosis Enumeration values: {benign, indeterminate, malignant}	enumeration
diagnosis_confirm_type	Method for diagnosing classification Enumeration values: {histopathology, single contributor clinical assessment, serial imaging showing no change, single image expert consensus, confocal microscopy}	enumeration
lesion_id	Secondary key: id for identifying lesion	string
age_approx	Age of the patient	integer
sex	Biological sex of patient Enumeration values: {male, female}	enumeration
anatom_site_general	General anatomic location Enumeration values: {head/neck, upper extremity, lower extremity, anterior torso, lateral torso, posterior torso, palms/soles, oral/genital}	enumeration
Fitzpatrick_skin_type	Fitzpatrick skin type Enumeration values: {I, II, III, IV, V, VI}	enumeration
image_data	Preprocessed dermoscopic image - Contains array of pixel values	array

hog_features	HOG features - Contains array of float numbers	array
pca_features	PCA features - Contains array of float numbers	array

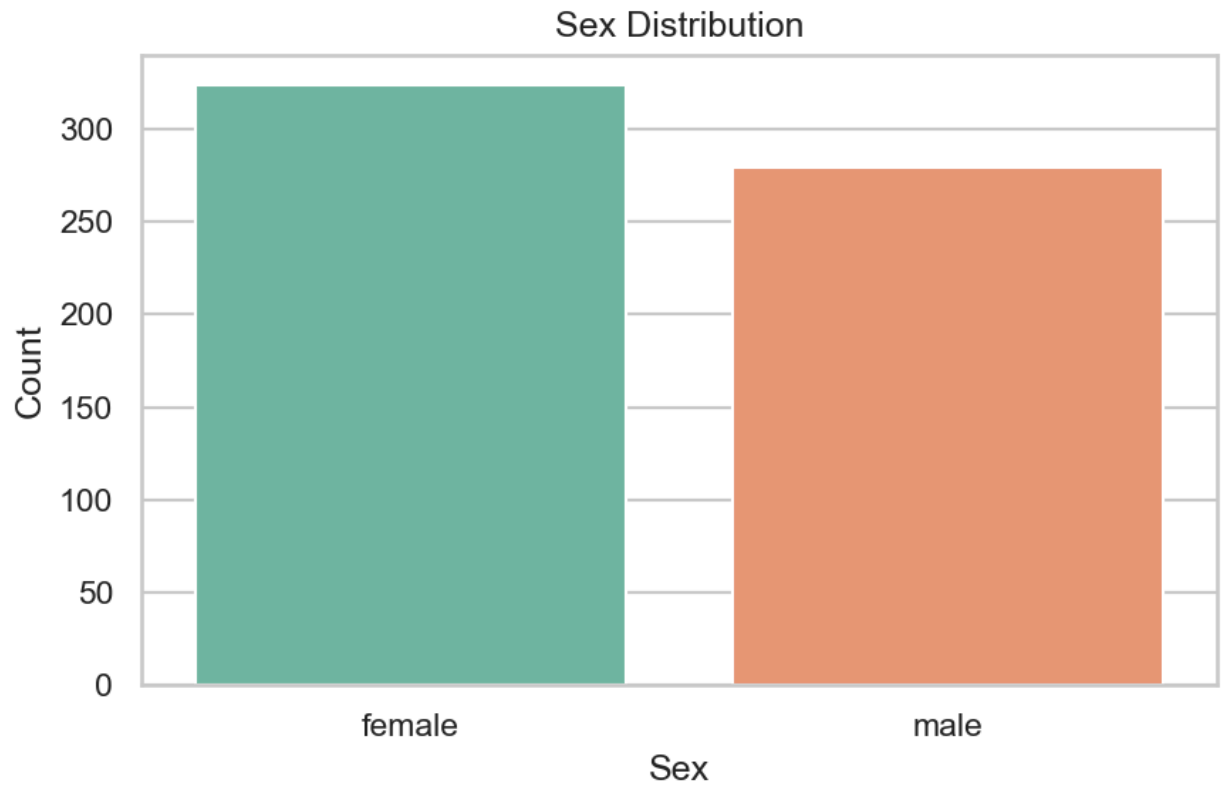
Descriptive Statistics

(1) Age Distribution
(2) Sex Distribution
(3) Lesion Anatomical Location Count
(4) Diagnosis Distribution
(5) Diagnosis Confirmation Count

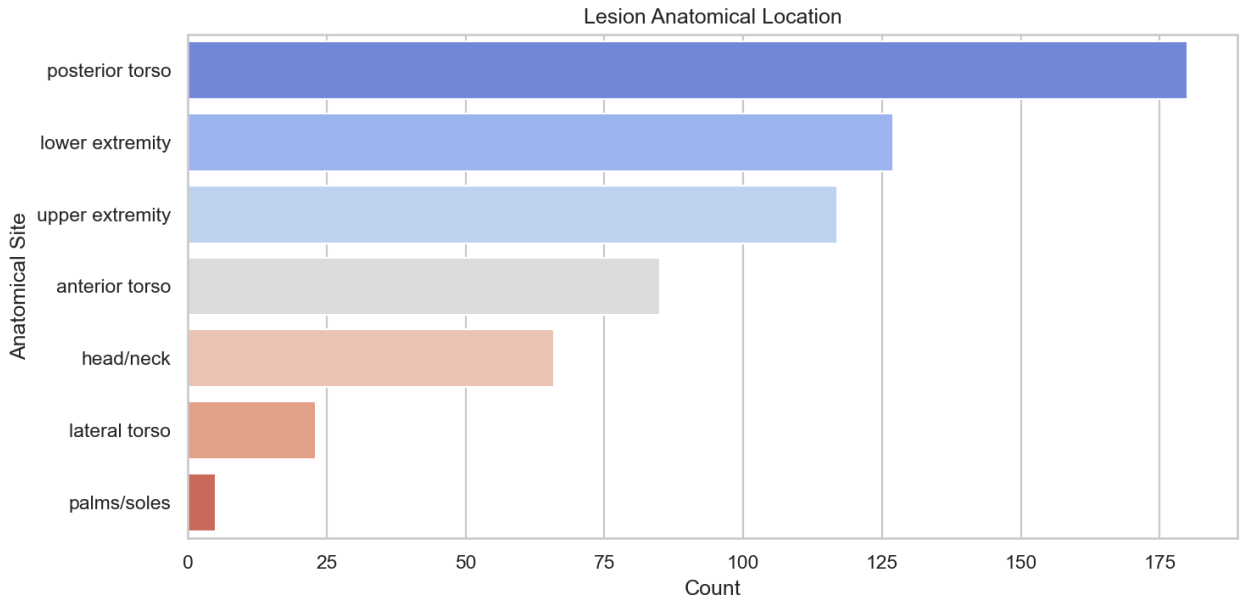
Age Distribution



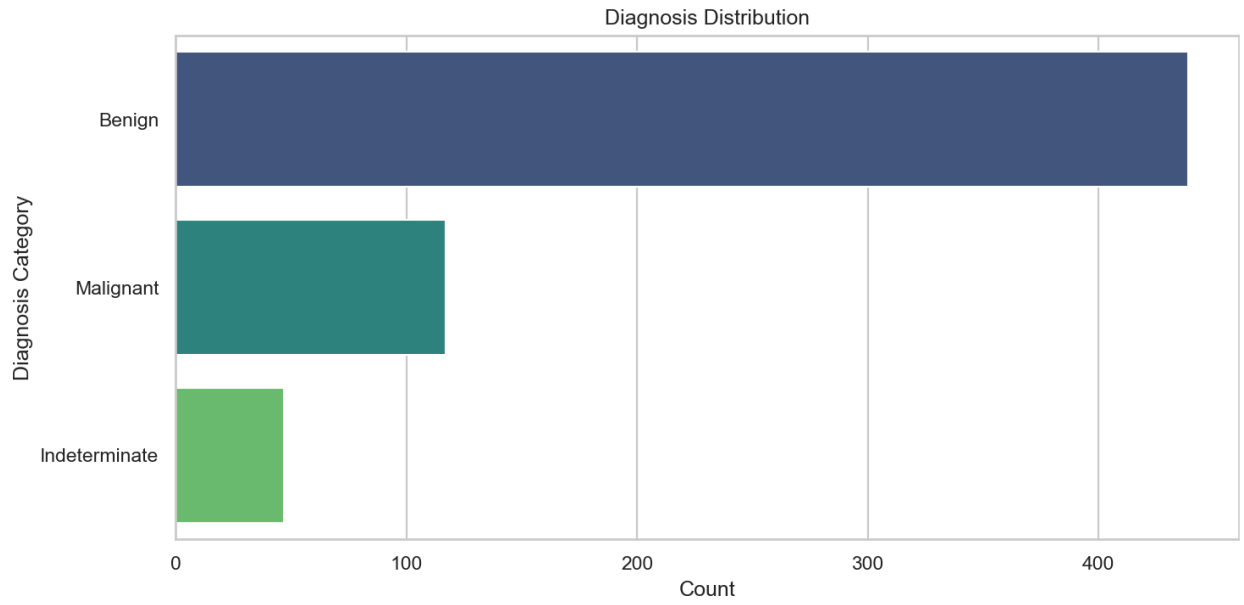
Sex Distribution



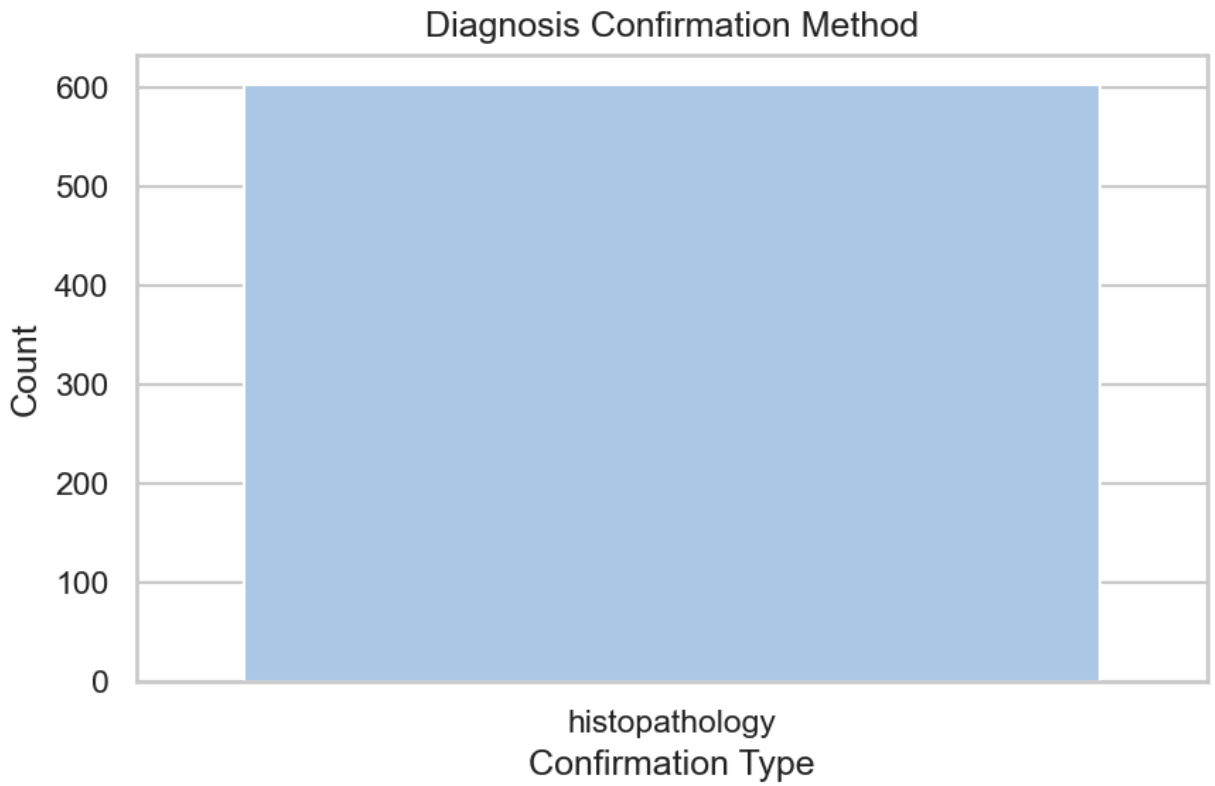
Lesion Anatomical Location Count



Diagnosis Distribution



Diagnosis Confirmation Count



The ISIC-Images.zip folder contains the following data:

Name	Information Obtained	Data Type
licenses	Contains file called CC-0.txt with license information	folder
attribution	Contains name of place where data was obtained from: ‘Memorial Sloan Kettering Cancer Center’	.txt file
ISIC_xxxxxxx	603 different images. Each image has a different 7 digit number that replaces the xxxxxxx	.jpg files
metadata.csv	same metadata csv as generated above	.csv file