Melanoma Disease Research: Analysis Based On Age, Gender, Skin Type, And Region For The United States

Gulhan BIZEL 🆔   Amey Gangadhar JAUNJARE 🆔   Virendra Mafaji RAJPUT 🆔

Dr.; Saint Peter's University, Frank J. Guarini School of Business, Data Science Institute, New Jersey, USA
Saint Peter's University, Frank J. Guarini School of Business, Data Science Institute, New Jersey, USA
Saint Peter's University, Frank J. Guarini School of Business, Data Science Institute, New Jersey, USA

ABSTRACT

Skin diseases are impacting people`s life like other diseases. Maybe, it doesn't hurt much so it takes a longer period of time to realise that it could be serious. Skin cancer which is also known as skin cancer is one of the most important skin diseases all around the world and in the United States, Its outcomes are at high risk for the human`s health. The most important risk factors are considered as genetic similarity and high exposure to the sunlight. The research reviews the dermatological diagnosis outcomes and patterns of melanoma in terms of age, gender, race, region of skin and geographical wise in the United States of America considering the regions by focusing on top 10 counties. According to the analysis done, the results reveal that melanoma is diagnosed more on white skin people than darker to black skin people. Gender wise, the analysis highlights that men are more like getting diagnosed more than women. In terms of age, it is mostly being observed between age of 60-65. Due to the weather nature of states, California and its counties are the top of the counties for the melanoma. The research describes the impact of age, gender, skin type, and regional wise trends of melanoma and it contributes to the literature about the patterns of melanoma and gives insights to healthcare companies, insurance companies and health centre's such as hospitals.

Keywords: Melanoma, dermatology, skin cancer

1. INTRODUCTION

In the skin of a human, the skin color is being developed by the presence of melanin cells. The darker the color of the skin may mean, skin has higher melanin cells and these cells protects the skin from sun. Once human skin is continuously exposed to ultraviolet it is regarded as the main factor of skin cancer (D'Orazio et al., 2013). Ionizing radiation, and pollutants are also related with skin cancers (El Ghissassi et al., 2009). The increased melanin appears to absorb UV light more efficiently (Cummins et al., 2006). The incidence rate of melanoma slowed to 3% per year from 1981 to 2000 (Surveillance, Epidemiology, and End Results Program, n.d.). The incidence of melanoma has been increasing for more than forty years in the US (Kopera, 2017). The melanoma incidence rate has continued to increase (Siegel et al., 2016). Melanoma is the only cancer that is not projected to meet federal objectives for reductions in death rates (Weir et al., 2015). Exposure to ultraviolet radiation increases melanoma risk (Armstrong & Kricker, 1993) (el Ghissassi et al., 2009b). Skin cancer is the most common type of cancer in the United States (*Melanoma Skin Cancer Statistics*, n.d.). The popular skin regions for melanoma on people of color are on non-sun-exposed skin such as the bottoms of the feet, the palms, and so on. (NCBI - WWW Error Blocked Diagnostic, n.d.).

Based on the literature survey it has been observed that the first aspect to melanoma is to sun exposure especially at the earliest time of a person`s life such as being as a child. The second possibility is getting exposed to high risk sun and in that case melanoma may occur on the several regions of the body such as arms. The society may not yet fully aware of the consequences of sun harm despite the communications observed by sun protecting brands and healthcare centers. The significance of this study may be to raise awareness on the risks of sun exposures while not disturbing much the lifestyle of the humans.

The research aims to analyse the effects of melanoma according to age, gender, skin type and region wise in the United States of America considering the 13 states and top 10 counties by looking at the diagnosis outcomes. This research describes what is known about the impact of age, gender, ethnicity, region of the skin and top counties for melanoma. Objective is the come to the conclusions about several aspects for the melanoma cases trends. So that, some suggestions may be advised for the society, researchers, healthcare centers, insurance companies and so on.

The research structure has been defined as the next section which is section 2 presents an explanation and overview of the methodology steps used in the research analysis. Section 3 presents the results of the data visualization by using Tableau data visualisation dashboard. The research describes the results by the usage of Tableau visualization dashboard as well as an analysis on the total cases of melanoma based on the skin type, gender, age, and regions affected by the disease. Finally, section 4 will be the conclusion for the research.

2. METHODOLOGY

This research is going to look for the results for melanoma which affected the different types of population in the United States of America based on age, gender, skin type and the region affected on the human body. By using several steps which are mentioned below for the analysis which are data profiling, data pre-processing, data extraction, and model classification in order as shown. The explanation for every step is mentioned as seen below in Figure 1.

Figure 1 Methodology process for the research.

Data profiling: The melanoma skin disease diagnosis data set which has been used in the research is extracted from the National Institutes of Health (NIH) for 13 states of the United States which are; Alaska, California, Connecticut, Georgia, Hawaii, Iowa, Kentucky, Louisiana, Michigan, New Jersey, New Mexico, Utah, and Washington. The data set is also drilling down into their counties.

The Surveillance, Epidemiology, and End Results (SEER) has been used for this research to accumulate cancer-related data to perform statistical analysis. The SEER SAT program was used in this research to make an analysis and to screen data regarding skin cancer.

Data pre-processing: This process helped to identify and correct the data quality. The melanoma cases dataset retrieved for the United States contains 200,472 columns and 19 rows. The dataset has 7 different variables which include patient ID, diagnostic confirmation, skin type record (fair skin/dark skin), gender, age at diagnosis, state-county, and region of skin cancer. The period of the dataset is between 2010 – 2016 years.

Data extraction: The next step following the data pre-processing is data extraction. It is about retrieving the data from the related source in a way that can be processed for further analytic purposes. The primary dataset is unstructured data so the first step is preparing the data by Microsoft Excel. Then, the next steps are importing the dataset in python and finding the unique variable. Data gathered from a source or a platform need to have data quality, accuracy in data, uniqueness, completeness, consistency, timelessness, and conformity. Uniqueness defines that the number of contents in the dataset such as rows and columns needs to have unique data and does not contain any unknown values which impact the accuracy in the data on the outcomes. Once the quality of data is assured, it is seen that the data is complete. Consistency is the second part of the analysis where it needs to find if there are findings such as null or any other unknown values in the dataset. Timelessness is the important aspect of data analysis where it defines the availability and accessibility of data to use further. Once the data has followed all the processes then it is confirmed that the data is ready for analysis. In the dataset patient ID is the unique character where the findings can be found. The last step is to re-arrange the data in some meaningful way so that it can be filtered out only the required information which helps to easily understand, visualize and analyse.

Model classification: Data analysis for this research was done by using Tableau software. It is a strong data visualization tool used in the business intelligence industry. By using a dashboard tool like Tableau, analysis becomes easier to understand rather than looking at a raw data. Tableau helps create data that can be understood by professionals (Reference Materials, n.d.). It gives opportunity to its users to build their own dashboards. Visualization was done according to the percentage of melanoma by skin type record and considering melanoma skin disease for top 10 counties in the U.S. by gender distribution between 2010-2016 years.

3. RESULTS

Melanoma is a disease which is not fading away in the new future for the population of the United States of America, possibly due to increasing ultraviolet radiation to previously protected regions. Therefore, the members examined melanoma incidence trends by age, gender, regions and conducted research analysis for the disease. The research also shows the comparison analysis incidence trends with the help of Tableau software. The results will be analysed in upcoming 4 different sections highlighting the correlations between the skin type, gender, age, regions, and counties for the population in the United States of America with the help of the analysis methodology such as data profiling, data pre-processing, data extraction, and model classification. The visualization of the results are performed by Tableau data visualisation software tool for better understanding.

3.1 Melanoma analysis outcomes by skin type

Before deep diving into the facts of the analysis, according to literature survey the epidermis of blacks has a sun-protection factor is higher versus the light skin (Kaidbey et al., 1979; Tadokoro et al., 2003). Higher amounts of melanin in people of the color filter at least twice as much UV radiation than the whites (Brenner and Hearing, 2008). This is the reason that the white population is the main victim (Gloster and Neal, 2006; Markovic et al., 2007). The data retrieved for this research is drilled down for the skin type which also refers to the ethnicity observation of the dataset.

Total Cases of Melanoma Depending On Skin Type

Race recode (W, B, AI, API)
■ American Indian/Alaska Native   ■ Asian or Pacific Islander   ■ Black   ■ Unknown   ■ White

Asian or Pacific Islander 1.56%
Black 1.64%
American Indian/Alaska Native 0.36%
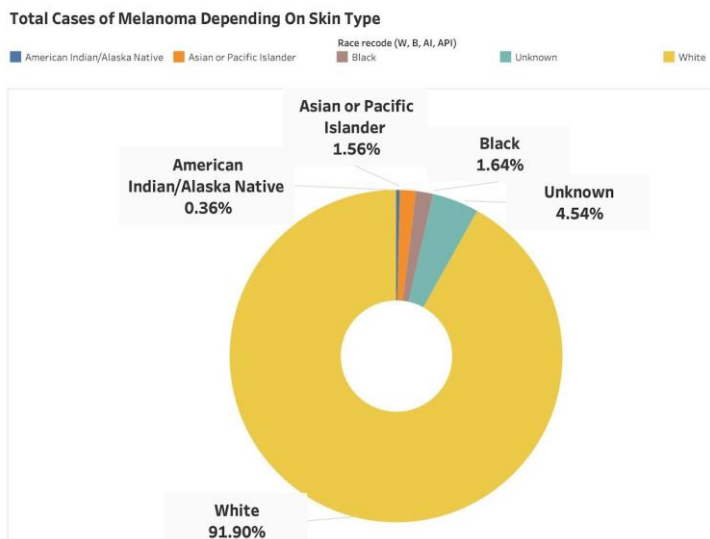Unknown 4.54%
White 91.90%

Figure 2 Total number of cases of melanoma considering the skin type

As shown in Figure 2 the total number of white skin melanoma cases in this research is 91.90% which is by far way more than black with 1.64%. The American Indian / Alaska Native population has been affected by a minority of 0.36%. The Asian or Pacific Islander population is affected by melanoma disease as well as and overall percentage of their affection rate is 1.56%. There is data available for the unknown population who are affected by melanoma which is 4.54%.

3.2 Melanoma analysis outcomes by gender by county

The data retrieved was available for 13 states and their counties. However, after cleaning and filtering the data, the top 10 counties were taken into consideration from the year 2010-2016 in order to focus on the most important ones. Hence, the below graph Figure 3 focuses majorly on the top 10 counties within the 3 states in the United States which are California, Utah and Washington.

Top 10 County In The U.S. by Gender Distribution Between 2010-2016 years

Sex   ■ Male   ■ Female

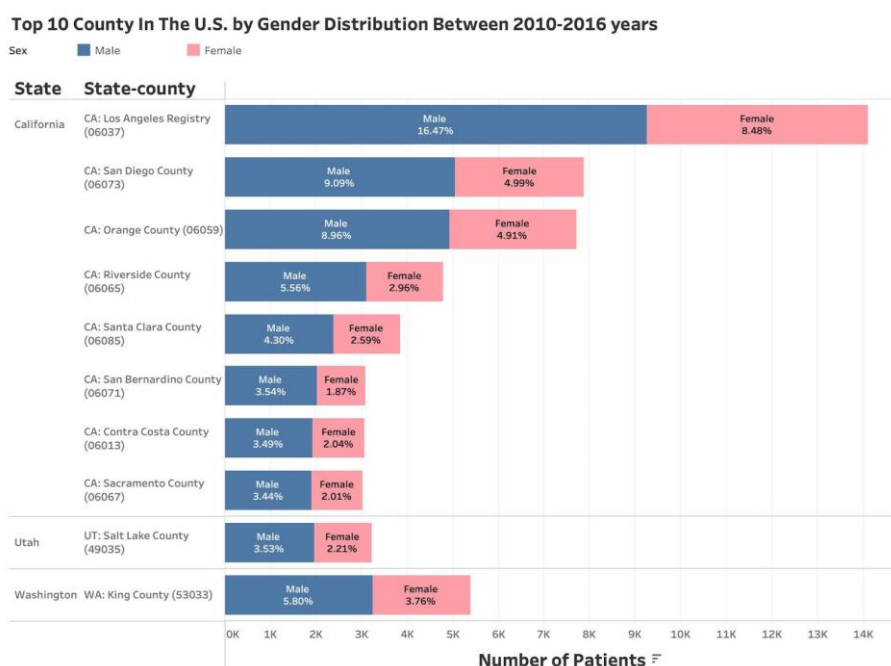| State	State-county		
California	CA: Los Angeles Registry (06037)	Male 16.47%	Female 8.48%
	CA: San Diego County (06073)	Male 9.09%	Female 4.99%
	CA: Orange County (06059)	Male 8.96%	Female 4.91%
	CA: Riverside County (06065)	Male 5.56%	Female 2.96%
	CA: Santa Clara County (06085)	Male 4.30%	Female 2.59%
	CA: San Bernardino County (06071)	Male 3.54%	Female 1.87%
	CA: Contra Costa County (06013)	Male 3.49%	Female 2.04%
	CA: Sacramento County (06067)	Male 3.44%	Female 2.01%
Utah	UT: Salt Lake County (49035)	Male 3.53%	Female 2.21%
Washington	WA: King County (53033)	Male 5.80%	Female 3.76%

Number of Patients

Figure 3 Melanoma comparison of gender in the U.S according to Top 10 counties based on the total number of patients

As shown in Figure 3, researched the states and their counties with a high UV index and distribution is based on male and female. Counties in California are topping the chart as seen above. The data was extracted between the

years 2010-2016. A large number of males have been affected with melanoma as compared to females. The highest cases of males reported were in California of 16.47%. The females affected were 8.48% in the same county. In Utah, Salt Lake county 3.53% of the male population has been affected compared to 2.21% of the women population as shown in Figure 3. Whereas, in the state of Washington, King county the count for impacted population is 5.80% for male compared to female with 3.76%.

3.3 Melanoma analysis outcomes by age and gender

The next step of the analysis is to understand the correlation between age and gender distribution on the patients who are diagnosed with melanoma.
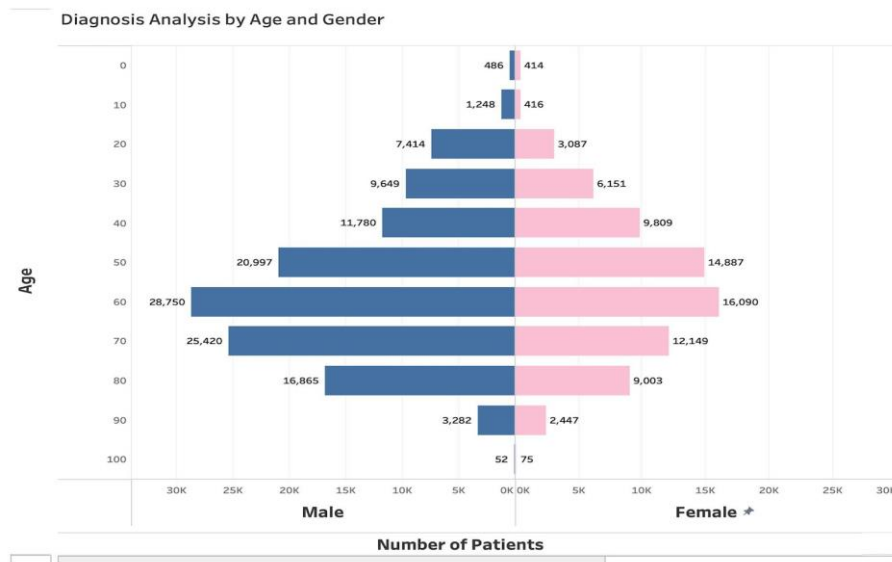

Figure 4 Melanoma analysis considering the number of patients based on age and gender

By age 65, making men 2 times as likely as women of the same age to get melanoma (American Academy of Dermatology | Public Education, n.d.). As shown in Figure 4, the highest peak recorded was in the male population of 28,750 for the age bracket of 60 to 65. Whereas in the female the highest peak is 16,090 for the same age group. The second maximum peak is 25,420 for the age bracket of 70 to 80 years old. While on the contrary, the second-highest peak is 14,887 for the age bracket of 45 to 55 for the female population. This may help to come to a conclusion that women are more protective than men and they may apply sunscreen more than men. Besides, cosmetic creams and makeup can be a factor on this considering women use makeup more than men. Another aspect can be considered as different skin structure between men and women.

3.4 Melanoma analysis outcomes by region of skin and skin colour type

As skin diseases, it is important to identify the region of the skin where melanoma is observed mostly. Besides, it is crucial to see if there is a patter with the colour of the skin.
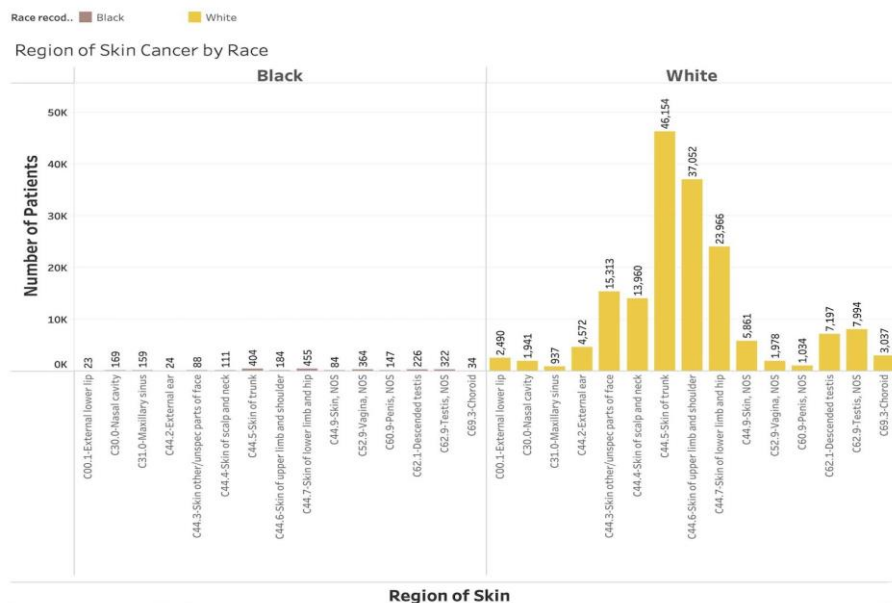

Figure 5 Melanoma analysis considering the number of patients categorized by region of skin and skin type

Figure 5 indicates the white skin type has mostly detected melanoma on the skin of the trunk (C44.5) and black skin type has mostly detected melanoma on the skin of the lower limb and hip (C44.7). Besides, it can be seen that the white population are the most affected people and the majorly affected region are the upper limb and shoulder areas with 46,154 patients. The majorly affected areas in males are the skin of upper limbs and shoulder followed by the skin of scalp and neck. The dataset retrieved for the analysis has provided the records like locational records, photographs, notes, charts, sketches, and descriptions of circumstances and materials. These are a part of systematic inspection, controlled recovery, or discovery of prehistoric or historic locations or things. The analysis of the site records from the International Classification of Diseases (ICD) indicates that the white skin type has mostly detected melanoma on the trunk (C44.5) which is the highest with 46,154 number of patients (C44.6) followed by the skin of the upper limb and shoulder with 37,052 number of patients (C44.7) and the skin of lower limb and hip with 23,966 as shown in the Figure 5. The darker the natural skin tone the more melanin it has. The highest peak of the black population is 455 number of patients. It is then followed by the skin of trunk with 404 and vagina with 364 cases respectively.

3.5 Melanoma analysis outcomes by gender and region of skin

Another analysis has done to understand the relationship between gender and the region of skin for melanoma disease to comment if there is a specific pattern related to it.

Figure 6 data visualisation is done the dataset extracted from SEER Primary site where Primary Label includes only the most important pieces of information about the chart. This includes the product name and contents. The black population has been affected in the areas of the lower limb and hip with 455 (C44.7) followed by the skin of the trunk with 404 patients (C44.5). Figure 5b shows the comparison by gender and their affected regions. The most affected region in females is the skin on the upper limb and shoulder (C44.6) with patients of 18,464 which is then followed by the skin of lower limb and hip (C44.7) and the skin of trunk (C44.5) with 17,677 and 16,932 respectively.
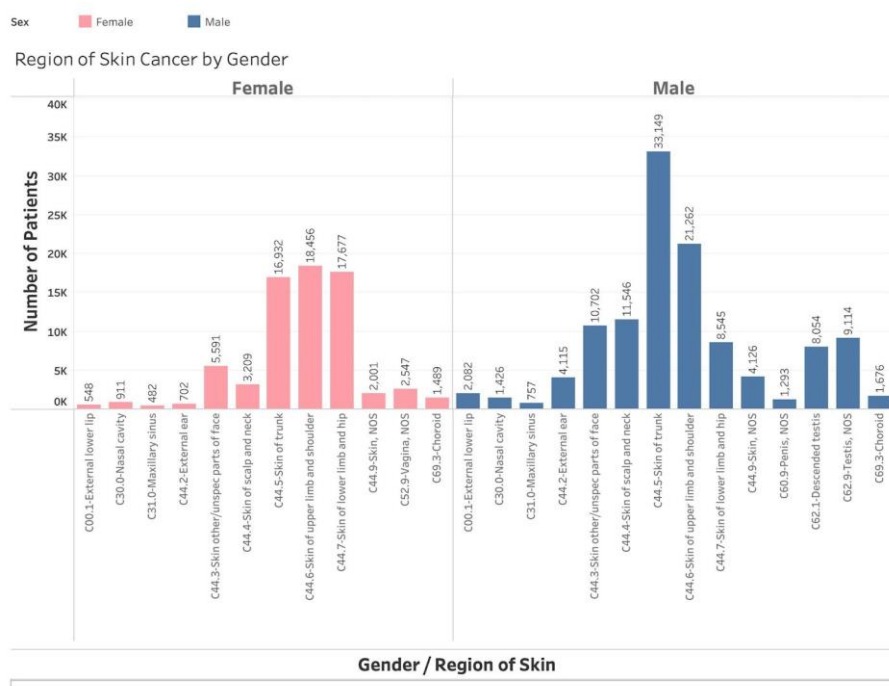

Figure 6 Melanoma analysis considering the number of patients categorized by gender and region of skin

4. CONCLUSION

This comprehensive study regarding the age, gender, skin type, and skin region trends in the incidence of melanoma in the United States of America showed that overall there are patterns for the incidence of the disease. Both genders have been effected due to the steadily increasing incidence in middle and old ages, with the steepest increase in males and older people. The analysis also demonstrates and emphasize the importance of prevention or elimination efforts, especially focusing on children, and their parents, to improve the UV radiation protection such as not to be exposed to sunlight much and secondary prevention activities to reduce the appearance of melanoma. All the research analysis suggests that this increase in the cases of melanomas of is observed on the specific regions of the skin such as the lower limb and hip along with the skin of the trunk. This leads to an outcome that lifestyle trends of people causing for UV radiation exposure in the recent decades. It may not be easy to educate populations for not sunbathing in the short term. However, these insights could be used for creating awareness for the population highlighting the regions of the skin that needs to be more protected from the dangerous UV of the sun. It may be

advised to use for screen protection for those areas of the body. Screen protection cosmetic brands may also put in their efforts into these directions. Instead of focusing on mostly women, awareness for the men population could be another factor to eliminate melanoma. Also, considering the outcomes of the study, middle-upper age level that has the highest risk could be target group of people for cosmetic brands but also for insurance companies. The main finding of the study is sun protection is still a threat for the society and humans have to be educated in the right path to prevent skin diseases and melanoma. Starting from childhood direct exposure to the sun must be prevents, especially direct exposure hours of the day, or more related months such as summer months and so on.

Regarding the state analysis for Utah and Washington could be a suggestion for future researches and these states may have more specific patterns than California in terms of weather conditions.

One of the obstacles for such a research study is the data collection process. The sources of data are limited. This research could be enriched if the data can be collected from many other databanks. Besides, the study was limited with several states due to shortage in dataset. As a solution, there should be a survey sent to the population regarding fair and dark skin so that it will be helpful in getting a clear picture for the development of the disease and the data would be more accurate and covering much wider geography. The research is showing a guidance tool for the future researches which can easily adapted with more recent numbers or for any other countries to make comparisons.

REFERENCES

American Academy of Dermatology | Public Education. (n.d.). American Academy of Dermatology | Public Education. Retrieved January 26, 2022, from https://www.aad.org/public

Armstrong, B. K., & Kricker, A. (1993). How much melanoma is caused by sun exposure? *Melanoma Research*, *3*(6), 395–402. https://doi.org/10.1097/00008390-199311000-00002

Brenner, M., & Hearing, V. J. (2007). The Protective Role of Melanin Against UV Damage in Human Skin†. *Photochemistry and Photobiology*, *84*(3), 539–549. https://doi.org/10.1111/j.1751-1097.2007.00226.x

Cummins, D. L., Cummins, J. M., Pantle, H., Silverman, M. A., Leonard, A. L., & Chanmugam, A. (2006). Cutaneous Malignant Melanoma. *Mayo Clinic Proceedings*, *81*(4), 500–507. https://doi.org/10.4065/81.4.500

D'Orazio, J., Jarrett, S., Amaro-Ortiz, A., & Scott, T. (2013). UV Radiation and the Skin. *International Journal of Molecular Sciences*, *14*(6), 12222–12248. https://doi.org/10.3390/ijms140612222

el Ghissassi, F., Baan, R., Straif, K., Grosse, Y., Secretan, B., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Freeman, C., Galichet, L., & Cogliano, V. (2009a). A review of human carcinogens—Part D: radiation. *The Lancet Oncology*, *10*(8), 751–752. https://doi.org/10.1016/s1470-2045(09)70213-x

Gloster, H. M., & Neal, K. (2006). Skin cancer in skin of color. *Journal of the American Academy of Dermatology*, *55*(5), 741–760. https://doi.org/10.1016/j.jaad.2005.08.063

Kaidbey, K. H., Agin, P. P., Sayre, R. M., & Kligman, A. M. (1979). Photoprotection by melanin—a comparison of black and Caucasian skin. *Journal of the American Academy of Dermatology*, *1*(3), 249–260. https://doi.org/10.1016/s0190-9622(79)70018-1

Kopera, D. (2017). Preventive treatment of sun-damaged skin: Topical treatment of actinic keratosis can prevent surgery in non-melanoma-skin-cancer. *Integrative Cancer Science and Therapeutics*, *4*(6). https://doi.org/10.15761/icst.1000261

Markovic, S. N., Erickson, L. A., Rao, R. D., Weenig, R. H., Pockaj, B. A., Bardia, A., Vachon, C. M., Schild, S. E., McWilliams, R. R., Hand, J. L., Laman, S. D., Kottschade, L. A., Maples, W. J., Pittelkow, M. R., Pulido, J. S., Cameron, J. D., & Creagan, E. T. (2007). Malignant Melanoma in the 21st Century, Part 2: Staging, Prognosis, and Treatment. *Mayo Clinic Proceedings*, *82*(4), 490–513. https://doi.org/10.4065/82.4.490

Melanoma Skin Cancer Statistics. (n.d.). American Cancer Society. https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html

NCBI - WWW Error Blocked Diagnostic. (n.d.). NCBI - WWW Error Blocked Diagnostic. Retrieved January 27, 2022, from https://pubmed.ncbi.nlm.nih.gov/25398793/

Reference materials. (n.d.). Tableau. https://www.tableau.com/resources/reference-materials

Siegel, R. L., Miller, K. D., & Jemal, A. (2016). Cancer statistics, 2016. *CA: A Cancer Journal for Clinicians*, *66*(1), 7–30. https://doi.org/10.3322/caac.21332

Surveillance, Epidemiology, and End Results Program. (n.d.). SEER. Retrieved January 11, 2022, from https://seer.cancer.gov

Tadokoro, Seiji & Shattil, Sanford & Eto, Koji & Tai, Vera & Liddington, Robert & Pereda, Jose & Ginsberg, Mark & Calderwood, David. (2003). Talin Binding to Integrin Tails: A Final Common Step in Integrin Activation. Science (New York, N.Y.). 302. 103-6. 10.1126/science.1086652.

Weir, H. K., Thompson, T. D., Soman, A., Møller, B., Leadbetter, S., & White, M. C. (2015). Meeting the Healthy People 2020 Objectives to Reduce Cancer Mortality. *Preventing Chronic Disease*, *12*. https://doi.org/10.5888/pcd12.140482