

Hackathon - Group 28

Gautam Reddy Chandupatla, Harsh Mahajan, Harshita Mishra

2023-02-08

Contents

1. Introduction
2. About the Dataset
3. Output
 - 3.1 Section 1 - Age and Lifestyle
 - 3.2 Section 2 - Sexually Transmitted Diseases (STDs)
 - 3.3 Section 3 - Medical Examinations
4. Conclusion
5. References

1. Introduction

Cervical cancer is cancer of the cells in the cervix, which is the lower part of the uterus and connects the vagina to the uterus. About 11,000 new cases of invasive cervical cancer are diagnosed each year in the U.S. However, the number of new cervical cancer cases has been declining steadily over the past decades. Although it is the most preventable type of cancer, each year cervical cancer kills about 4,000 women in the U.S. and about 300,000 women worldwide. In this report, we will discuss about some major risk factors that lead to Cervical Cancer.

2. About the Dataset

This dataset is obtained from UCI repository which contains a list of risk factors for Cervical Cancer leading to a Biopsy Examination. This dataset contains 858 rows and 36 columns.

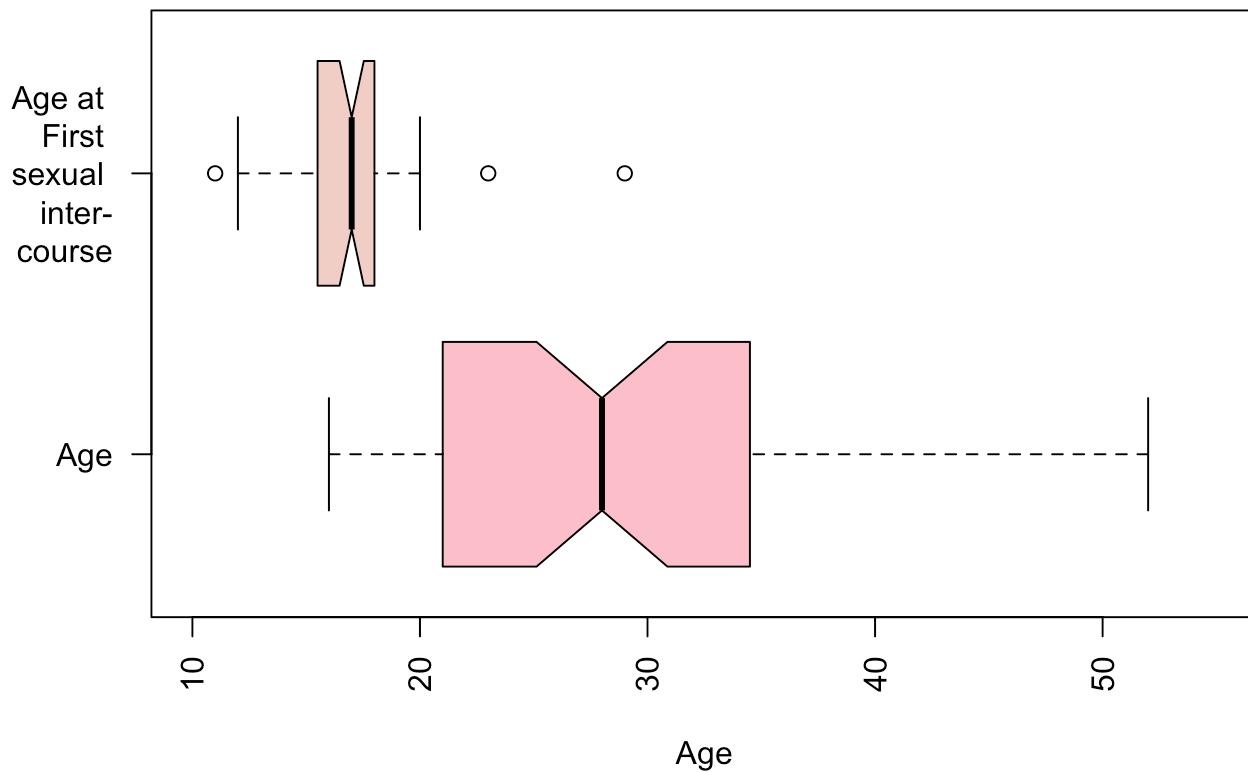
3. Output

3.1 Section - 1: Age and Lifestyle

Is "AGE" of a women a risk factor for cervical cancer?

We will be analyzing the data for patients who have higher risk of cervical cancer. Here the data is filtered based on our requirement and the following boxplots are generated.

Boxplot of Women diagnosed with Cervical Cancer



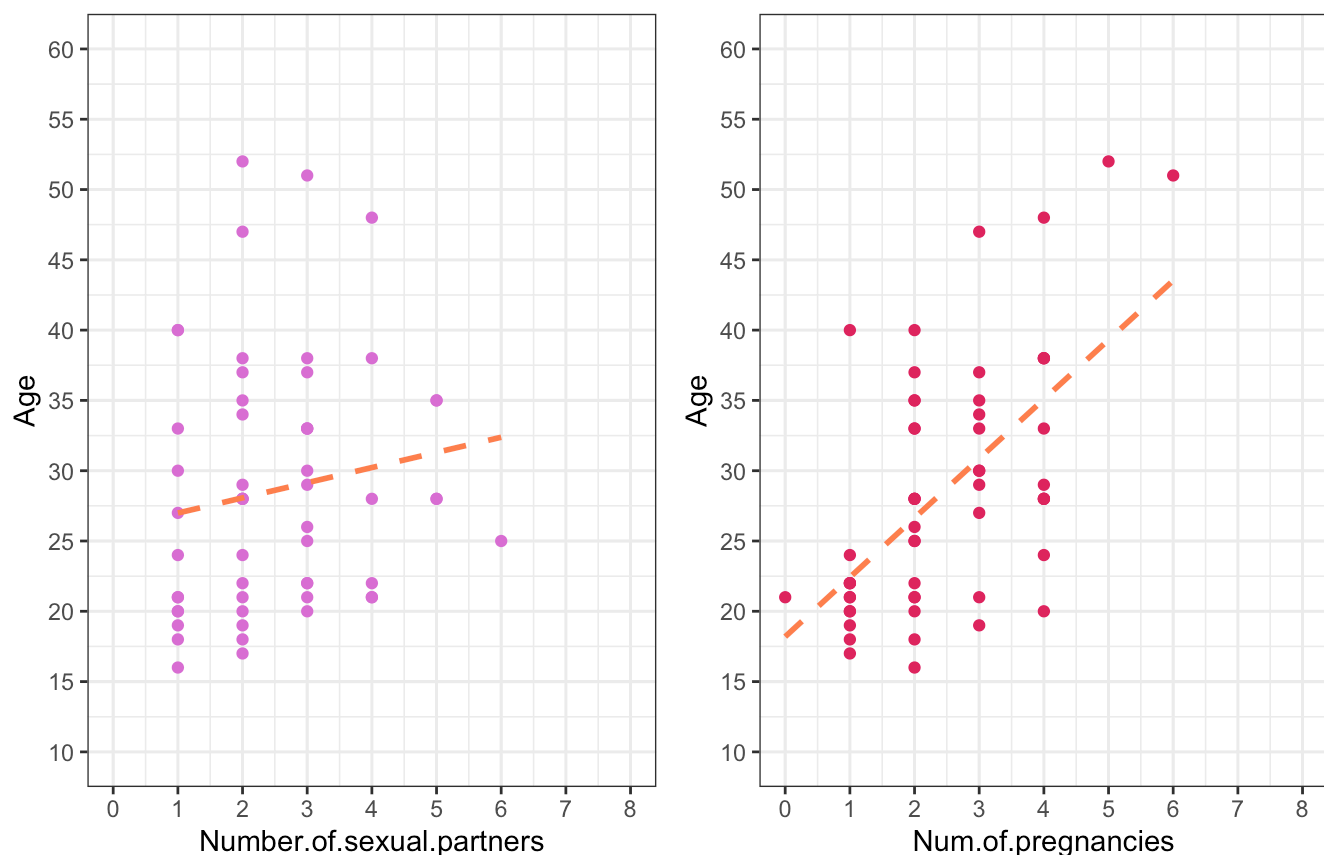
From this boxplot, the median age of women having first sexual intercourse is at 17 years old, are twice as likely to get cervical cancer later than women who didn't have their first intercourse until 20 years or older. This may be because many women fail to receive the sexual health information, education, and access to the care they need to lead healthy lives. The risk of cervical cancer can be reduced by giving awareness to women about sex education. There are some outliers, and it is difficult to predict the reason as we don't know the complete medical history of the patient.

Secondly, here we can notice that the risk of Cervical cancer is most often diagnosed between the ages of 21 and 34, with the median age of affected women being 28 years. About 25% of women are prone to be diagnosed after the age 35. These cases usually occur in people who did not receive regular screening tests. The reason for low screening tests may be due to high poverty levels. Other factors such as lack of health insurance, limited transportation, and language difficulties also effects a poor woman's access to screening services.

Is there a risk involved in having more sexual partners and pregnancies?

Our analysis here is confined to patients who have a risk of cervical cancer.

Scattered plots showing number of sexual partners and number of pregnancies along with age of women having risk of cervical cancer



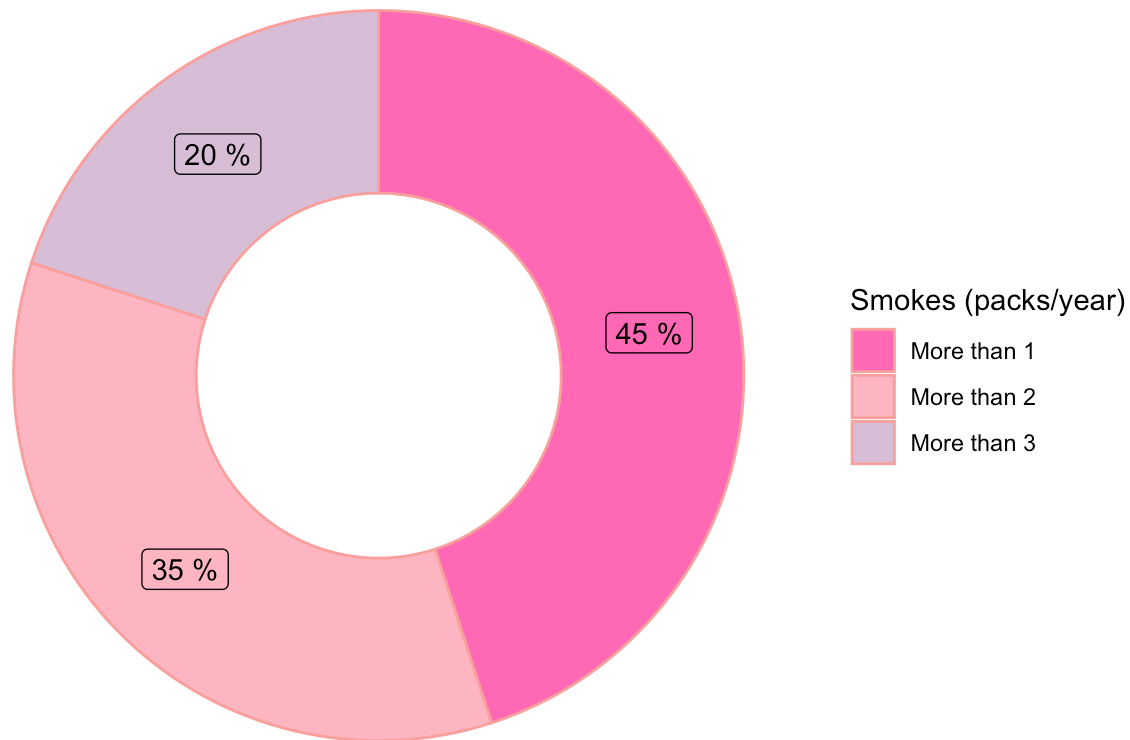
We have maintained same scale in both plots in order to simplify our visualization.

From this plot we can conclude that women who have had three or more pregnancies are more likely to develop cervical cancer. This is probably due to the increased exposure to HPV infection with sexual activity. Also hormonal changes during pregnancy is possibly making women more susceptible to HPV infection or cancer growth. Another thought is that pregnant women might have weaker immune systems, allowing for HPV infection and cancer growth.

Former graph shows median number of sexual partners is 3. Women who have had three or more sexual partners are more likely to develop cervical cancer. In case of higher sexual partners there is a higher risk of having HPV and other STDs. These diseases play a vital roles in causing cervical cancer, which will be discussed in detailed in section-2.

Does smoking causes risk of cervical cancer?

Risk of Smoking on Cervical Cancer



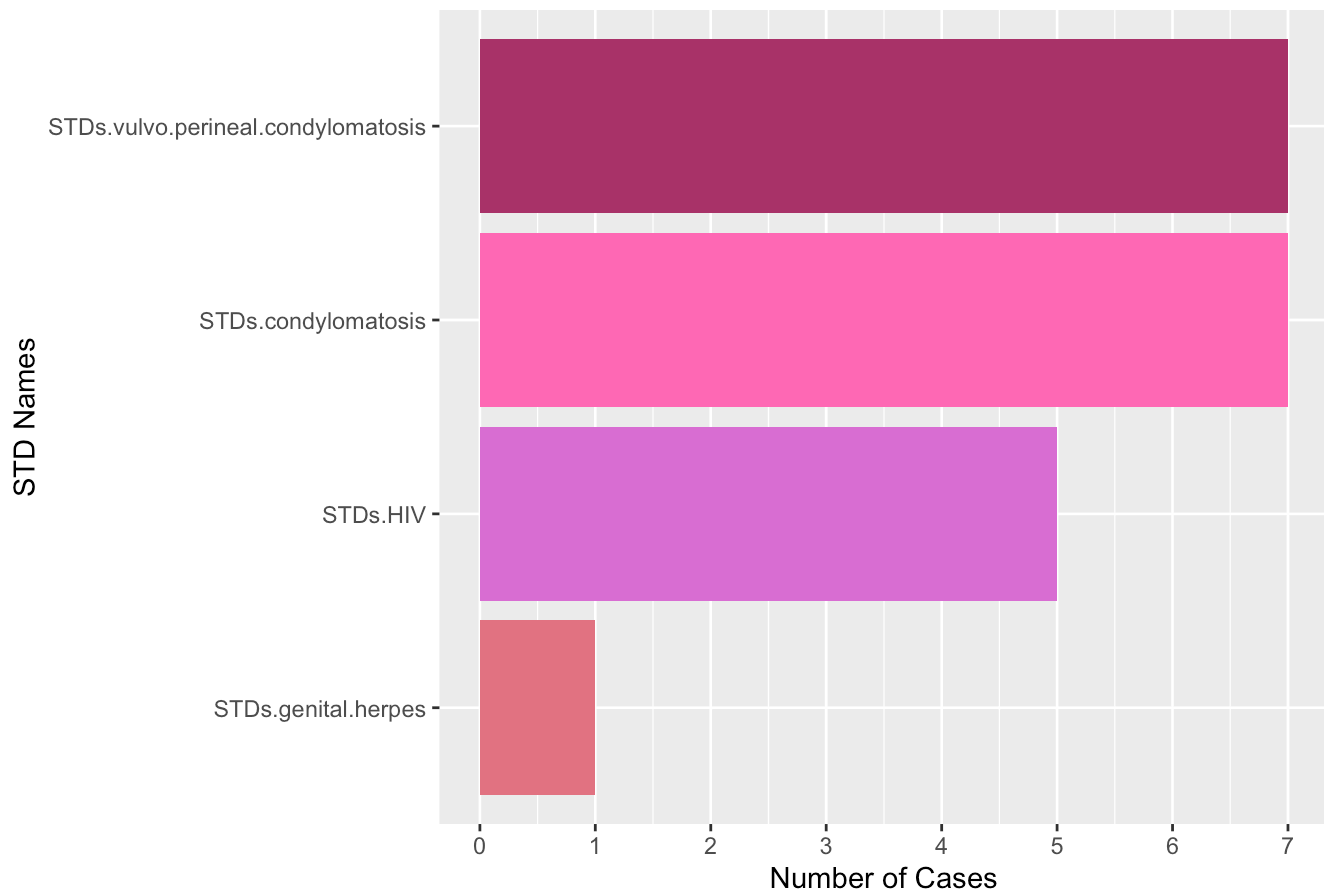
The above donut chart is plotted for patients who have risk of cervical cancer, it can be observed that 45% women smoke more than 1 pack/year, 35% smoke more than 2 packs/year, 20% smoke more than 3 packs/year. Here the number of smoke packs increases the risk of cervical cancer. This is because smoking allows cancer causing chemicals to be absorbed through the lungs and carried in the blood stream throughout the body. Tobacco by-products can be found in the cervical mucus of women who smoke.

Apart from cervical cancer, smoking also increases the risk of developing other diseases like lung cancer, lung disease, heart disease and other major health issues. It is never easy to quit smoking, especially if you have been a smoker for a long time and it has become part of your routine. So if a women is smoking, then she is advised to have her screening tests done regularly.

Section 2 - Sexually Transmitted Diseases (STDs)

Which STDs increase chance of cervical cancer?

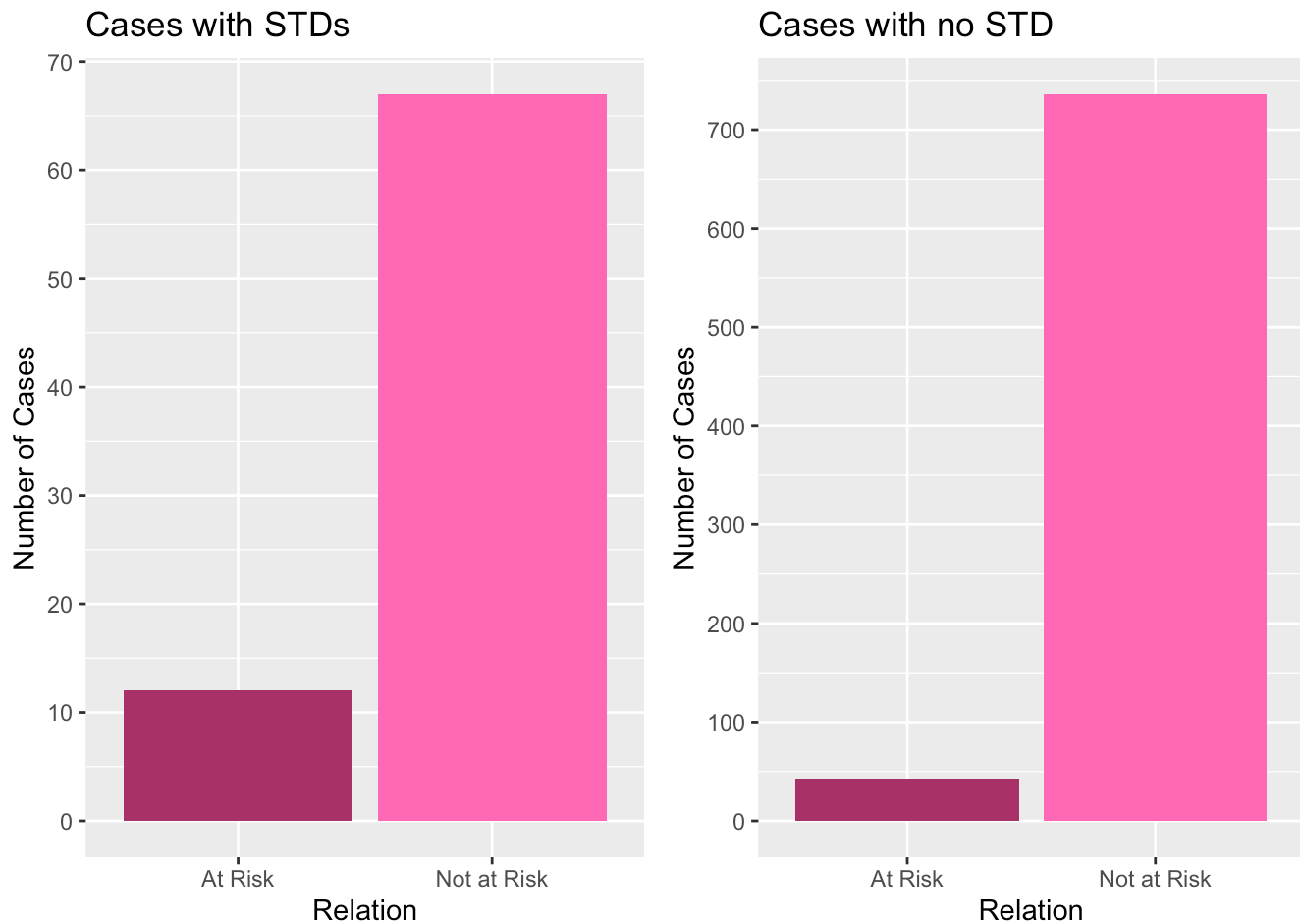
Most common STDs found in Cervical Cancer patients



The visualization above shows the most common STDs found in patients at risk of cervical cancer. Vulvo Perineal Condylomatosis and condylomatosis are both types of the HPV virus (Human papillomavirus). The HPV virus has been linked strongly with the risk of cervical cancer. As a domino effect, having multiple sexual partners has also been identified as a key factor in at risk patients.

Having a compromised immune system is also a factor hence 3rd on the list is the HIV virus (human immunodeficiency virus). As the name suggests, this virus makes the patient immuno-compromised, hence making vulnerable to several infections one of which can be HPV or other STDs.

Can STDs cause cervical cancer?

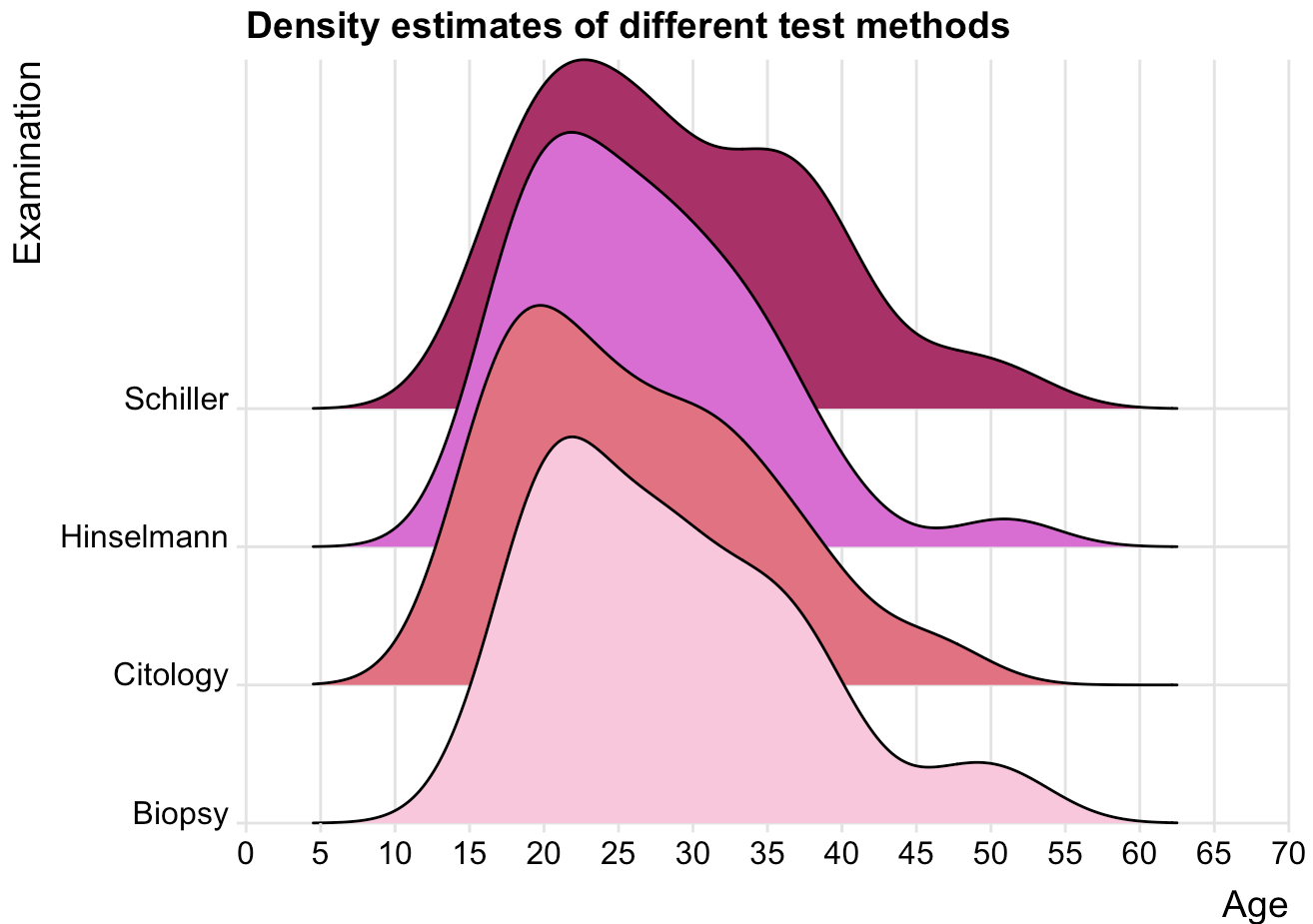


The above graphs are a comparative study about the impact of having STDs and being at risk of cervical cancer. In the plot displaying the distribution of patients with STDs we can see that the rate of At Risk patients is much higher than the plot that shows the distribution of patients without STDs. To quantify this comparison: 15.2 percent of people with an STD are at risk of Cervical cancer. On the contrary, only 5.5 percent of people without an STD are at risk of cervical cancer. Having an STD makes it almost 3 times more likely to be at risk of getting cervical cancer, making STDs a major contributing factor for at risk patients.

Section-3: Medical Examinations

How accurate are the prediction results of different medical examinations? Do they differ?

The medical examinations used to diagnose Cervical Cancer are Schiller, Hinselmann, Citology and Biopsy.



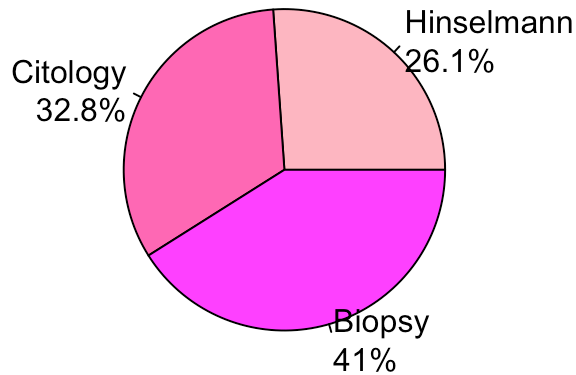
Density plots of different medical examinations are generated for patients having risk of cervical cancer. All these show variations between the age of 17-23 years. As we have discussed in Section-1, having first sexual intercourse at age below 20 years is one of the factors.

It is clear that there is not much variation between these plots, as they have different medical procedures to declare results. Here the most common method for detecting cervical cancer is by taking “Biopsy” test of the tissue. However other tests are not too bad either.

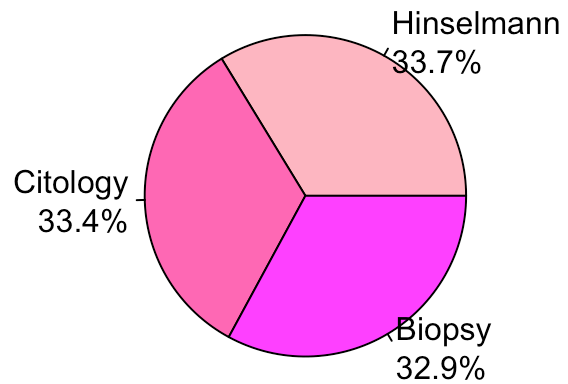
However women may get just one medical test or a combination of these depending on the recommendations of the doctors. The best medical examination to adopt depends on several other factors, including what stage of cancer a women has.

What are the most effective tests for detecting cervical cancer?

Positive Cases



Negative Cases



The Hinselmann, Citology, and Biopsy tests are all important medical tests used to screen for various types of cancer, including cervical cancer. The pie charts help to illustrate the number of positive and negative cases for these tests.

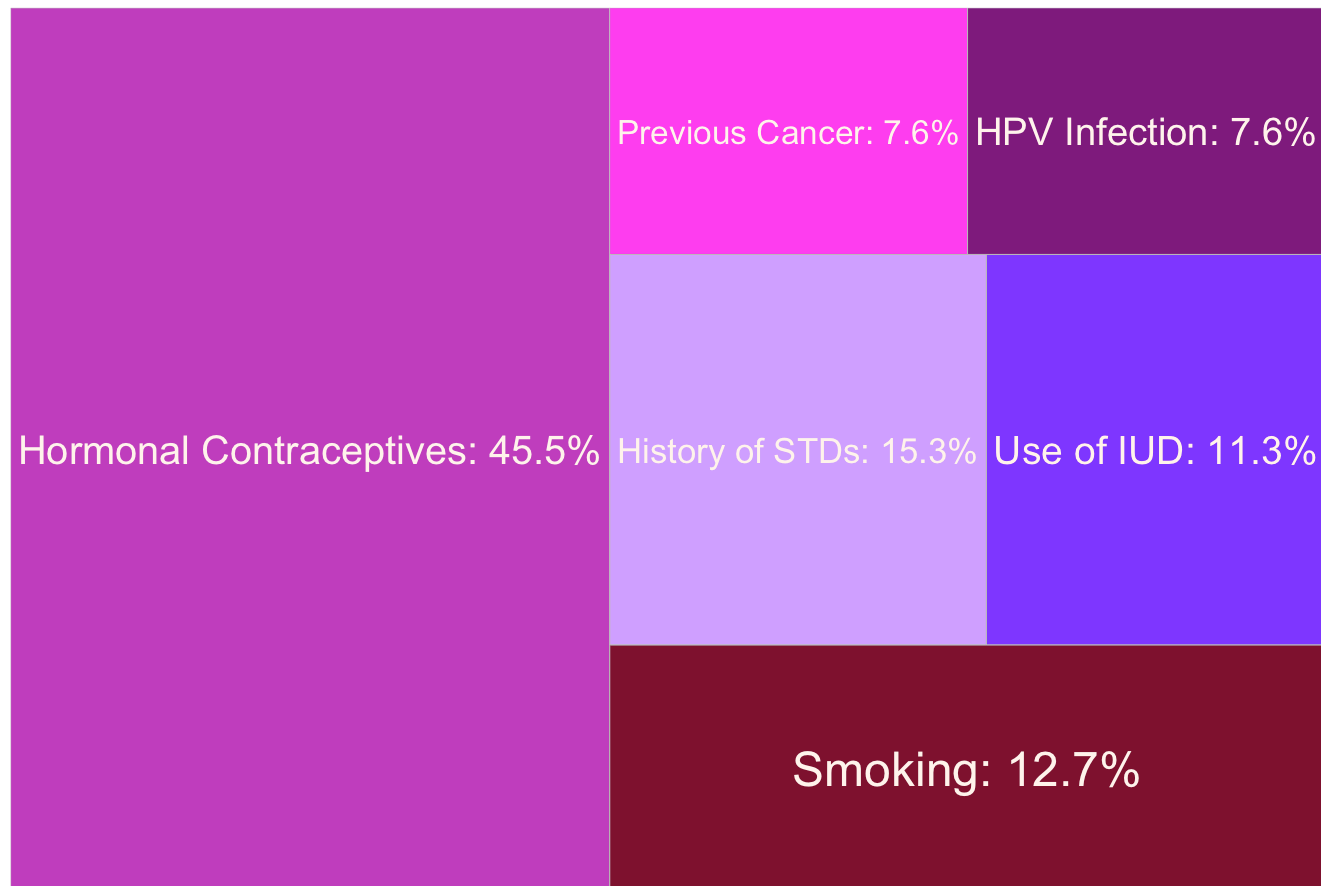
In general, a high number of negative cases is a positive outcome for the individual being tested, as it suggests that there are no signs of cancer or abnormal cells. However, a positive result can indicate the presence of abnormal cells or other signs of cancer, which may require further testing or treatment.

The proportion of positive cases in the Biopsy test is relatively higher than the Hinselmann and Citology tests, which may indicate that the Biopsy test is more effective at identifying cancerous or pre-cancerous cells.

Overall, the pie charts can help medical professionals and researchers to understand the distribution of positive and negative cases for these important medical tests, which can guide decision-making and further research.

4. Conclusion:

Most common conditions found in At-Risk Patients



Cervical Cancer has several risk factors and they've been categorized and analyzed in this report. The above treemap serves the purpose of concluding this report by summarizing and quantifying the most common risk factors that have been observed in patients at risk of cervical cancer. Hormonal contraceptives being the most common factor found between at risk patients, taking these pills for more than 5-10 years exponentially increases the risk of cervical cancer. Followed by Smoking, use of IUDs and having a history of STDs. Despite the dimensional restrictions of this dataset, the conclusions drawn are well in-line with the prevalence of the subject. There are a variety of tests available for detecting cervical cancer with Biopsy being the best method. However, as we studied in section 3, a person can pick any test that they feel comfortable with as long as they get tested and diagnose cancer in it's early stages to seek proper treatment.

5. References:

-> <https://www.kaggle.com/datasets/loveall/cervical-cancer-risk-classification>
(<https://www.kaggle.com/datasets/loveall/cervical-cancer-risk-classification>)

-> <https://lovelacecancercenter.com/blog/risk-factors-cervical-cancer#>
(<https://lovelacecancercenter.com/blog/risk-factors-cervical-cancer#>):~:text=Multiple%20pregnancies,until%2025%20years%20or%20older.

-> <https://www.cancer.org/cancer/cervical-cancer/causes-risks-prevention/risk-factors.html>
(<https://www.cancer.org/cancer/cervical-cancer/causes-risks-prevention/risk-factors.html>)

-> <https://www.healthywomen.org/condition/stages-of-cervical-cancer>
(<https://www.healthywomen.org/condition/stages-of-cervical-cancer>)

