İstanbul Gedik University

Faculty of Engineering

Course: MCT202 Statistics and Probability

Task: Assignment – I

Hüsnü Berke KAYA

191023001

17/04/2021

Instructor's Name and Surname: Prof. Dr. Ali Cevat Taşıran

PLACE OF BAD HABİTS AMONG

OTHER EFFECTS OF CANCER

Hüsnü Berke Kaya

17/04/2021

ABSTARCT

Bad habits like alcoholism, smoking dependency and drugs are one of the most fatal problems of the society. People use this kind of products to numb themselves or just because they are used to these stuff. There is so many reasons that people are in this psychological situation. The most common reason is to evade their problems in their lives. The short feeling of happiness while using this products seems easier to people than solving their problems. Another big reason is wannabe. Most young people takes this habits from persons who they are inspired from. The reason of this being a problem is not just addiction. These things also results in big medical problems. Drugs can end in deformation in tissue and organs, alcoholism can cause cirrhosis, smoking can result in over-increased heart rates. Whereas, in this article the common problem of this three will be discussed, which is cancer.

Keywords: Cancer, alcohol, smoking, drugs, addiction, bad habits

1. Introduction

Cancer is one of the biggest medical issue of 21st century. There is no certain cure for it and people are still not aware about it. The cancer can be defined as non-controlled over-multiplication of cells in body caused by deformation in DNAs and RNAs. Lots of things can cause cancer. There are various reasons of cancer. Fossil gases, radiation, depression are some of them. In this study the most common reason will be mentioned.

Despite it is confirmed again and again by experts a part of society still thinks bad habits like smoking, alcoholism, taking drugs etc. doesn't result in cancer.

Although these are not the only reasons for cancer it is known that their over-usage tend to result in cancer.

Aim of this article is to show the effects of these bad habits on cancer patients. This will be done through analyzing a dataset of cancer patients and whether they have these bad habits or how often they take these stuff.

2. Theoretical Framework

In the article "Environmental and occupational causes of cancer: A call to act on what we know" [1] (Richard W. Clapp, Genevieve K. Howe, Molly M. Jacobs) it is suggested that %80 percent of the reason for cancer are environmental causes. This article provides and focuses on a table with environmental and occupational reasons for cancer.

Table 1
Sources and uses of environmental and occupational carcinogens

Category	Carcinogenic agent	Source/uses	Strong[a]	Suspected[a]
Aromatic amines	Benzidine, 2-naphylamine, 4,4′-methylenebis 2-choloraniline (MOCA), chlornaphazine heterocyclic aromatic amines	Used as antioxidants in the production of rubber and cutting oils, as intermediates in azo dye manufacturing, and as pesticides. Common contaminant in chemical and mechanic industries and aluminum transformation and an air contaminant from tobacco smoking. Used widely in the textile and beautician (as hair dyes) industries.	Bladder (Benzidine, 2-naphylamine, 4,4′-methylenebis 2-choloraniline (MOCA), chlornaphazine)	Prostate (heterocyclic aromatic amines)
Chlorination by-products	Trihalomethanes	Trihalomethanes include chloroform, bromodichloromethane, chlorodibromomethane, and bromoform. Result from the interaction of chlorine with organic chemicals. Several halogenated compounds may form from these reactions although trihalomethanes are the most common. Brominated by-products are also formed from the reaction of chlorinated by-products with low levels of bromide in drinking water.		Bladder; rectal
Environmental tobacco smoke	Contains more than 50 known carcinogens	Environmental tobacco smoke (ETS), also considered passive smoke is a combination of smoke emitted from the burning end of a cigarette, cigar, or pipe, and smoke exhaled by the smoker.	Lung; breast	
Metals	Arsenic	Is produced commercially as a by-product of nonferrous metal production, primarily from copper production, comprising greater than 10% of dust content in some smelter operations. Inorganic arsenic is primarily used to preserve wood, but is also used as a pesticide mainly on cotton plants.	Bladder; lung; skin; soft tissue sarcoma (angio-sarcoma of the liver)	Brain/CNS; renal; liver and biliary; prostate; soft tissue sarcoma
	Beryllium	Used in the nuclear, aircraft and medical devices industry. Used also as an alloy or in specialty ceramics for electrical and electronic applications. Found as a contaminant in the combustion of coal and fuel oil.	Lung	
	Cadmium	Occurs naturally in ores together with zinc, lead and copper. Used as stabilizers in PVC products, color pigment, several alloys and now most commonly in re-chargeable nickel−cadmium batteries. Also present as a pollutant in phosphate fertilizers.	Lung	Pancreatic; renal; prostate
	Chromium	Chromium is used in steel and other alloy production. Chromium III and Chromium VI are used in chrome plating, the manufacture of dyes and pigments, leather tanning and wood preserving.	Lung; nasal and nasopharynx	
	Lead	Used primarily in the production of batteries, ammunition, metal products such as solder and pipers and devices to shield X-rays. Lead is also found in gasoline, paints, ceramic products, caulking, and pipe solder, but has been reduced dramatically in the US.		Brain/CNS; lead; renal; stomach
	Mercury	Used to produce chlorine gas and caustic soda. Also used in thermometers, dental fillings, and batteries.		Brain/CNS
	Nickel	Used primarily as an alloy in stainless steel. Also used in nickel plating and battery production.	Lung; nasal and nasopharynx	Laryngeal; pancreatic; stomach
Metalworking fluids and/or mineral oils	Straight oils, soluble oils, synthetic and semi-synthetic fluids	Used in a variety of industries including metal machining, print press operating and cotton and jute spinning.	Bladder; laryngeal; lung nasal and nasopharynx (mineral oils); rectal; skin; stomach	Esophageal; pancreatic; prostate

|				
Natural fibers/dust	Asbestos	An inorganic naturally occurring fibrous silicate particle used primarily in acoustical and thermal insulation. Asbestos fibers can be divided into two groups: chrysotile (most widely used) and amphibole which include amosite, crocidolite, anthophyllite, actinolite and tremolite fibers.	Laryngeal; lung; mesothelioma	
	Silica	An inorganic particle used in foundries, brick making and sandblasting.	Lung	
	Talc containing asbestiform fibers	A mineral used in the manufacture of pottery, paper, paint and cosmetics	Lung	
	Wood	Used primarily in carpentry, joinery and in furniture and cabinetry making	Lung; nasal and nasopharynx	Laryngeal
Pesticides	Herbicides, fungicides and insecticides [note: for specific pesticides see Clapp et al. 2005 [14]	Used for preventing, destroying, repelling or mitigating any pest or in use as a plant regulator, defoliant or desiccant. The majority of pesticides as registered with the US. EPA are used in agricultural applications, although residential application is also an important source.		Brain/CNS; breast; Hogkin's; leukemia; lung; multiple myeloma; NHL; ovarian; pancreatic; renal; soft tissue sarcoma; stomach; testicular
Petrochemicals and combustion by-products	Petroleum products, motor vehicle exhaust (including diesel), polycyclic aromatic hydrocarbons (PAHs), soot, and dioxins	Petrochemicals are derived from natural gas or petroleum and used to produce a variety of other chemicals and materials including pesticides, plastics, medicines and dyes. Substances can be produced as the building blocks for other products, but mainly result from the incomplete combustion of burning coal, oil, gas (diesel exhaust), household waste, tobacco and other organic substances. Dioxins are a class of chemical that are the by-products of combustion processes containing chlorine and carbon-based chemicals such as polyvinyl chloride (PVC) plastics. Dioxins are also created during the chlorine-bleaching processes for whitening paper and wood pulp.	Lung (PAHs, air pollution including diesel exhaust, soot, dioxin); NHL (dioxin); soft tissue sarcoma (dioxin); skin (PAHs)	Bladder (PAHs); breast (dioxin); esophageal (soot); multiple myeloma (dioxin); prostate (dioxin and PAHs)
Radiation	Ionizing radiation	Any one of several types of particles and rays given off by radioactive material, high-voltage equipment, nuclear reactions and stars. Alpha and beta particles, X-rays and gamma rays are radiation particles of concern to human health.	Bone; brain and central nervous system; breast; leukemia; liver and biliary; lung; multiple myeloma; soft tissue sarcoma; skin; thyroid	Bladder; colon; nasal and nasopharynx; ovarian; stomach
	Non-ionizing	Comprised of microwaves and electromagnetic frequencies including radio waves and extremely low-frequency electromagnetic fields.		Brain; breast; leukemia
	Ultraviolet radiation	Ultraviolet radiation is part of the solar radiation emitted by the sun.	Melanoma	

Table 1 (*continued*)

| Category	Carcinogenic agent	Source/uses	Strong[a]	Suspected[a]
Reactive chemicals	Butadiene	Used in the production of polymers for the manufacture of styrene-butadiene rubber for tires, nitrile rubber for hoses, gaskets, adhesives and footwear; acrylonitrile-butadiene-styrene polymers for parts, pipes, and various appliances; and styrene−butadiene latexes for paints and carpet backing.		Leukemia
	Ethylene oxide	Used as a sterilant, disinfectant and pesticide. It is also used as a raw ingredient in making resins, films and antifreeze.	Leukemia	Breast
	Formaldehyde	Used primarily in the production of urea, phenol or melamine resins for molded products such as appliances, electric controls, and telephones; in particle-board and plywood and in surface coatings.		Nasal and nasopharynx
	Mustard gas	Produced and used primarily in World War I as a chemical warfare agent.	Lung	Laryngeal
	Sulfuric acid	Used widely in industry for the production of isopropanol, ethanol; treatment of metals; and the manufacture of soaps, detergents and batteries.	Laryngeal	Lung
	Vinyl chloride	Vinyl chloride is used in polyvinyl resins for the production of plastic pipes, floor coverings, and in electrical and transportation applications.	Liver and biliary; soft tissue sarcoma (angio-sarcoma of the liver)	
Solvents	Benzene	Used as an intermediate in the production of plastics, resins and some synthetic and nylon fibers. Also used to make some types of rubbers, lubricants, dyes, detergents, drugs and pesticides. Is also found in crude oil, gasoline and cigarette smoke.	Leukemia; NHL	Brain/CNS; lung; nasal and nasopharynx; multiple myeloma
	Carbon tetrachloride	Used primarily in various industrial applications. Before being banned, was also used in the production of refrigeration fluid and propellants for aerosol cans, as a pesticide, as a cleaning fluid and degreasing agent, in fire extinguishers, and in spot removers.		Leukemia
	Styrene	Used in the production of rubber, plastic, insulation, fiberglass, pipes, automobile parts, food containers and carpet backing.		NHL
	Methylene chloride	Used primarily as a solvent in a variety of industrial applications and as a paint strippers. It may also be found in some aerosol and pesticide products and in the production of photographic film.		Brain/CNS; liver and biliary
	Toluene	Used in the production of paints, paint thinners, fingernail polish, lacquers, adhesives and rubber. Also used in some printing and leather tanning processes.		Brain/CNS; lung; rectal
	Trichloroethylene (TCE)	Used mainly for degreasing metal parts. Previous used as a dry-cleaning agent. TCE may be found in printing inks, varnishes, adhesives, paints and lacquers. Important contaminant in the general environment as a result of emissions & leakage from industrial settings.	Liver and biliary; renal	Cervical; Hodgkin's; leukemia; NHL; renal
	Tetrachloroethylene (PCE)	Used to degrease metal parts and as a solvent in a variety of industrial applications. Since 1930s used by an increasingly large percentage of US. dry-cleaning operations.		Bladder; cervical; esophageal; NHL; renal
	Xylene(s)	Used as a cleaning agent, a thinner for paint and in paint and varnishes. Used in printing rubber and leather industries and found in small amounts in gasoline and airplane fuel.		Brain/CNS; rectal

BBC's article "Non-communicable diseases" is an article about diseases like cancer and diabetes. In this article reasons of cancer is slightly touched on and its focus is on human fault with respect to graphs and tables below:

This table of data and bar chart show the number of deaths that could be avoided in England and Wales in 2013 from different causes.

Cause of death	Number of deaths that could be avoided, rate per 100 000 of population
Cancer	80
Cardiovascular disease	62
Drug-related deaths	11
Infections	3
Injury	35
Respiratory diseases	23
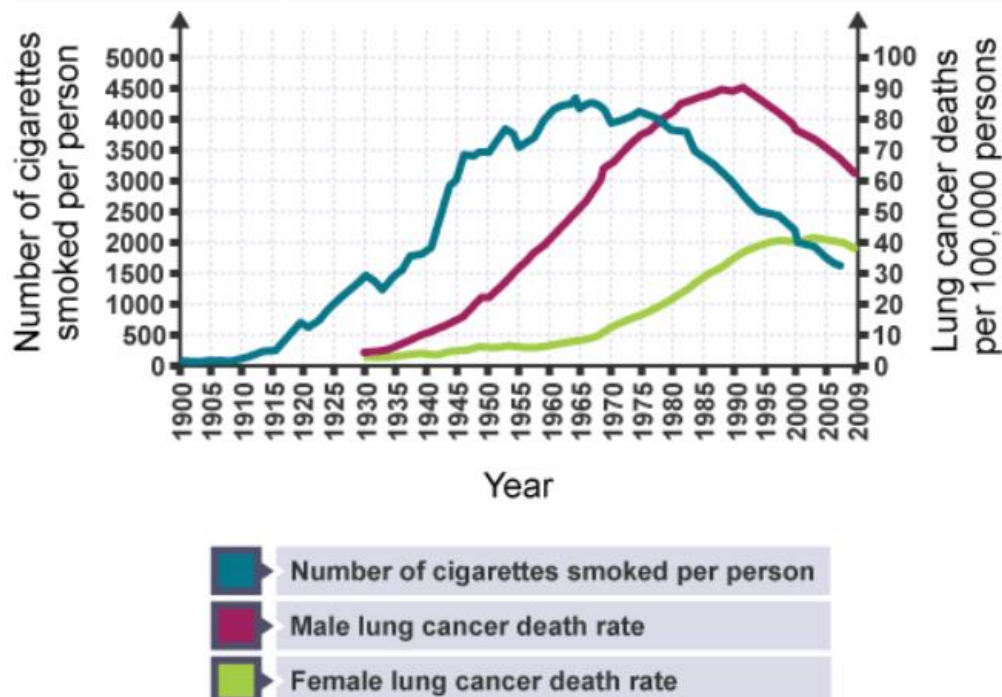
AMERICAN CANCER SOCIETY

Deaths that could be avoided (rate per 100,000 of population) vs **Cause of death**
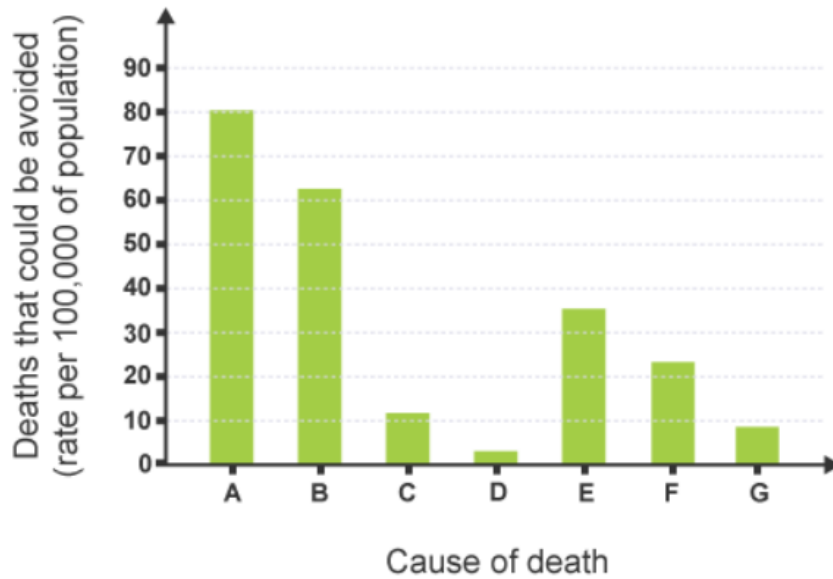
A	Cancer	F	Respiratory diseases
B	Cardiovascular disease	G	Other causes
C	Drug-related deaths		
D	Infections		
E	Injury		

This **stacked bar chart** shows proportions of cases of cancer in the USA connected with the Human papilloma virus (HPV), and those where no connection has been established.

Annual number of cases in USA vs **Type of cancer**
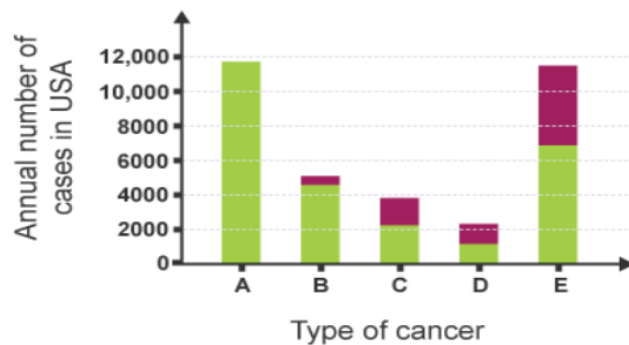
A	Cervical		HPV cases
B	Anus		Non-HPV cases
C	Vaginal and vulval		
D	Penile		
E	Mouth and throat		

Source: Centers for Disease Control and Prevention

To summarize the theoretical framework part of the article it can be said that there is previous studies about the case "causes of cancer" that focuses on either environment- like the article "**Environmental and occupational causes of cancer: A call to act on what we know**" -or human fault -like BBC's orientation on "**Non-communicable diseases**"-. In this article an orientation similar to BBC's is going to be taken because this article is about bad habits like smoking and alcoholism's effects on cancer.

3.Data and Variables

CLASSIFICATION	VARIABLE	DESCRPITION	SUMMARY
PATİENT INFO	Patient ID	Identification of the patient	P1-P1000
	Age	Age of the patient	Min:14 Max:73 Mean:37.17
	Gender	Gender of the patient(1for male and 2 for female)	Male: 598 Patients Female: 402 Patients
Human Fault	Alcohol Use	Level of alcohol usage of the patient	Min: 1 Max:8 Mean:4.563
	Balanced Diet	Quality of the patient's diet(8 for good, 1 for bad)	Min:1 Max:8 Mean:4.491

|			
Human Fault	Smoking	Level of active smoking of the patient	Min:1 Max:8 Mean:3.948
	Passive Smoking	Level of passive smoking of the patient	Min:1 Max:8 Mean:4.195
Other Reasons	Air pollution	Level of air pollution patient was under influence of	Min:1 Max:8 Mean:3.384
	Dust allergy	Level of dust intolerance of the patient	Min:1 Max:8 Mean:5.165
	Occupational Hazards	Level of hazards experienced because of the patient's job	Min:1 Max:7 Mean:4.84
	Genetic Risk	Level of risk inherited from parents of the patient	Min:1 Max:8 Mean:4.58
	Chronic Lung Disease	Level of other diseases of the patient in lung that is occurring more than 1 year	Min:1 Max:7 Mean:4.38
Complaints	Chest Pain	Level of chest pain of the patient	Min:1 Max:9 Mean:4.438

	Coughing Of Blood	Level of blood coughing of the patient	Min:1 Max:9 Mean:4.859
	Fatigue	Level of tiredness of the patient	Min:1 Max:9 Mean:3.856
	Weight Loss	Level of weight loss of the patient	Min:1 Max:8 Mean:3.855
	Shortness Of Breath	Level of the breathing difficulty of the patient	Min:1 Max:9 Mean:4.24
	Wheezing	Level of the wheezing of the patient	Min:1 Max:8 Mean:3.777
Complaints	Swallowing difficulty	Level of difficulty on swallowing of patient	Min:1 Max:8 Mean:3.746
	Clubbing of Finger Nails	Level of finger nail clubbing of the patient	Min:1 Max:9 Mean:3.923
	Frequent Cold	Amount of cold occurred to patient frequently to the disease	Min:1 Max:7 Mean:5.736

|	Dry Cough	Level of dryness in patients cough	Min:1 Max:7 Mean:3.853
Complaints	Snoring	Level of snoring of the patient	Min:1 Max:7 Mean:2.926
Level of the disease	Level	Level of the cancer in patients body	Low:303 patients Medium:332 patients High:365 patients

Statistical modelling
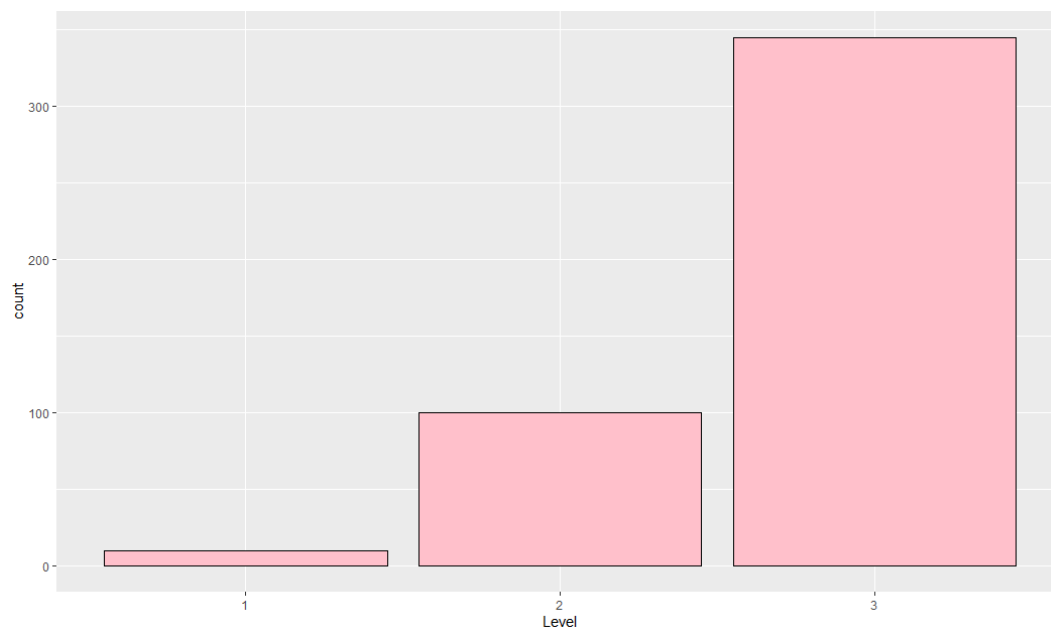
Graph (1): Level of illness on patients those are affected heavily by environment
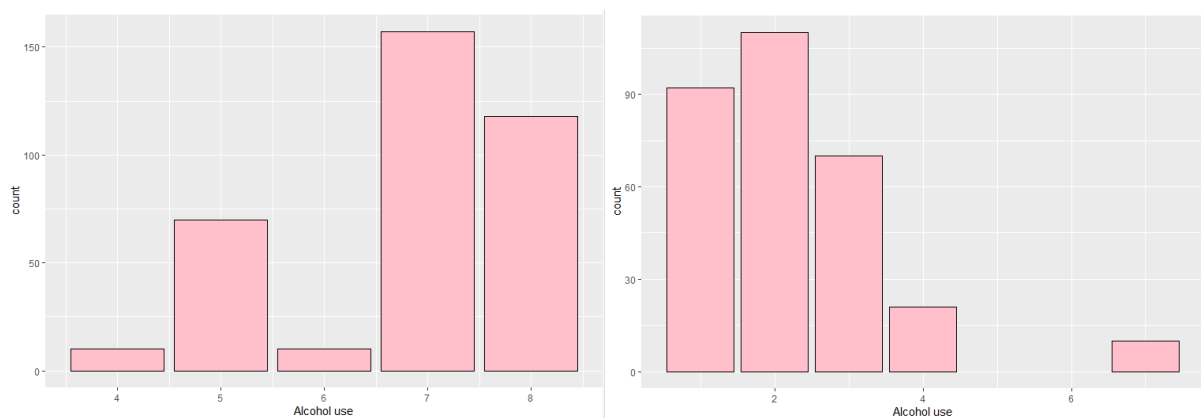
Graph (2) and Graph (3): Alcohol usage of patients with low (3) and high (2) level of cancer
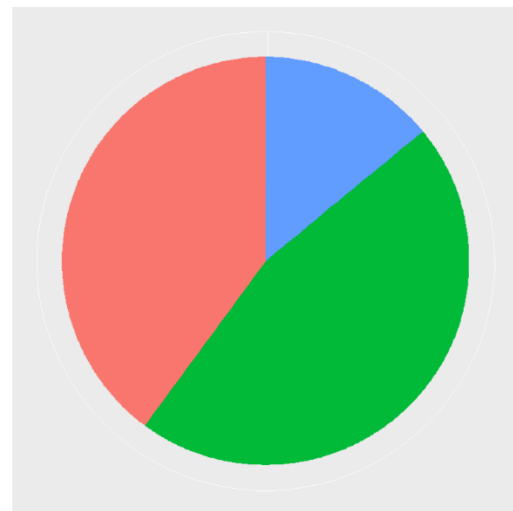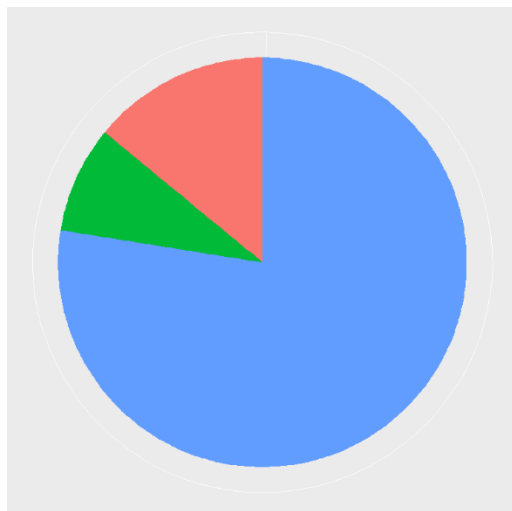
Graph (4) and Graph (5): Cancer level of patients with light smoking (5) and heavy smoking (4)
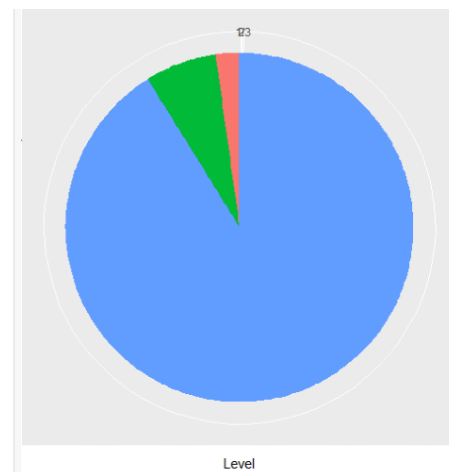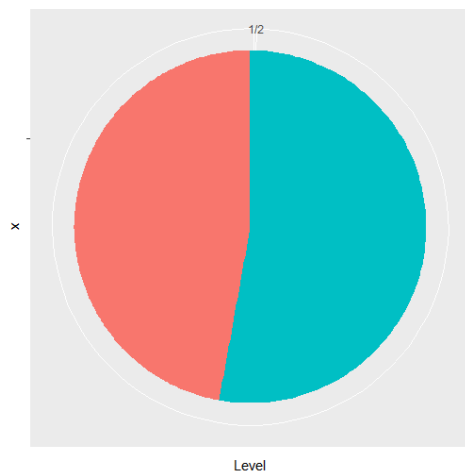
Graph (6) and Graph (7): Comparison of patients with high level of bad habits (7) and without bad habits (6)

```
 Patient Id              Age            Gender        Air Pollution   Alcohol use      Dust Allergy    OccuPational Hazards
 Length:1000      Min.    :14.00    Min.   :1.000    Min.   :1.00    Min.   :1.000    Min.   :1.000   Min.    :1.00
 Class :character 1st Qu.:27.75    1st Qu.:1.000    1st Qu.:2.00    1st Qu.:2.000    1st Qu.:4.000   1st Qu.:3.00
 Mode  :character Median :36.00    Median :1.000    Median :3.00    Median :5.000    Median :6.000   Median :5.00
                  Mean    :37.17    Mean   :1.402    Mean   :3.84    Mean   :4.563    Mean   :5.165   Mean    :4.84
                  3rd Qu.:45.00    3rd Qu.:2.000    3rd Qu.:6.00    3rd Qu.:7.000    3rd Qu.:7.000   3rd Qu.:7.00
                  Max.    :73.00    Max.   :2.000    Max.   :8.00    Max.   :8.000    Max.   :8.000   Max.    :8.00
  Genetic Risk   chronic Lung Disease Balanced Diet     Obesity         Smoking        Passive Smoker    Chest Pain
 Min.   :1.00    Min.    :1.00        Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.    :1.000
 1st Qu.:2.00    1st Qu.:3.00         1st Qu.:2.000   1st Qu.:3.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
 Median :5.00    Median :4.00         Median :4.000   Median :4.000   Median :3.000   Median :4.000   Median :4.000
 Mean   :4.58    Mean    :4.38        Mean   :4.491   Mean   :4.465   Mean   :3.948   Mean   :4.195   Mean    :4.438
 3rd Qu.:7.00    3rd Qu.:6.00         3rd Qu.:7.000   3rd Qu.:7.000   3rd Qu.:7.000   3rd Qu.:7.000   3rd Qu.:7.000
 Max.   :7.00    Max.    :7.00        Max.   :7.000   Max.   :7.000   Max.   :8.000   Max.   :8.000   Max.    :9.000
 Coughing of Blood   Fatigue         Weight Loss     Shortness of Breath   Wheezing     Swallowing Difficulty
 Min.   :1.000    Min.    :1.000    Min.   :1.000    Min.   :1.00      Min.   :1.000   Min.    :1.000
 1st Qu.:3.000    1st Qu.:2.000    1st Qu.:2.000    1st Qu.:2.00      1st Qu.:2.000   1st Qu.:2.000
 Median :4.000    Median :3.000    Median :3.000    Median :4.00      Median :4.000   Median :4.000
 Mean   :4.859    Mean    :3.856    Mean   :3.855    Mean   :4.24      Mean   :3.777   Mean    :3.746
 3rd Qu.:7.000    3rd Qu.:5.000    3rd Qu.:6.000    3rd Qu.:6.00      3rd Qu.:5.000   3rd Qu.:5.000
 Max.   :9.000    Max.    :9.000    Max.   :8.000    Max.   :9.00      Max.   :8.000   Max.    :8.000
 Clubbing of Finger Nails Frequent Cold   Dry Cough       Snoring          Level
 Min.   :1.000          Min.    :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
 1st Qu.:3.000          1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:1.000
 Median :4.000          Median :3.000   Median :4.000   Median :3.000   Median :2.000
 Mean   :3.923          Mean    :3.536   Mean   :3.853   Mean   :2.926   Mean   :2.062
 3rd Qu.:5.000          3rd Qu.:5.000   3rd Qu.:6.000   3rd Qu.:4.000   3rd Qu.:3.000
 Max.   :9.000          Max.    :7.000   Max.   :7.000   Max.   :7.000   Max.   :3.000
```

Result

As stated at the starting of the article, drinking alcohol, smoking, drugs etc. can cause cancer. This can be understood by comparing subsets of patients with addiction to this matters (more than level 5) have more objects compared to those without addiction(level 5 or less than 5). By comparing graph (6) and graph (7) it can also be seen that patients with these bad habits tend do have higher levels of cancer. Another analysis we can do is that people with no bad habits did not have cancer level higher than "medium". Another comparison that can be done is heavy smokers have a 78% distribution of level 3 cancer.

By summarizing all these analyses and more from the dataset, data of patients with addictions can be compared to patients who don't. Other comparisons that can be done with this dataset can be: environmental causes' effects on cancer level, different complaints of patients those are in effect of other causes like genetics, environmental hazards, air pollution etc.

Conclusion

By comparing and analysing these data with its subsets cancer patients tend to have addictions to these materials. Whereas, this analysis proves that although environmental and genetic causes have less effect on having the sickness, they affect the patients' sickness level by increasing it. In conclusion of these two analysis it can be said that bad habits tend to have more effect whether having the disease or not having it, whereas environmental causes heavily effect patients' level of illness.

In comparison of graph (2) and graph (3) a statement can be made as "Patients with heavy alcohol usage have more probability to have cancer compared to non-heavy alcohol drinkers." As stated in the article alcohol causes cancer for a variety of reasons:

1- The ethanol in alcoholic drinks break down to acetaldehyde, a known carcinogen that damages DNA by stopping cells from repairing damage which results in growth of cancer cells.

2-Alcohol provides empty calories which leads to weight gain that increases risk of cancer.

3-Alcohol prevents body from breaking down some vitamins and nutrients that protects body from cancer.[1]

---

[1]"7 thing to know about cancer" by Danielle Underferth: Link to article

By analysing graph(4) and graph(5) it can be seen that light smokers have a very low percentage of having high level of cancer and heavy smokers have a very high chance of having higher levels of cancer. 65% percent of heavy smokers have a high level of cancer.

Another comparison that can be done is between percentages of patients with high level of cancer within heavy alcohol drinkers and heavy smokers. Approximately 78% of the high level cancer patients are heavy alcohol drinkers and approximately 75% of high level cancer patients are heavy smokers. This comparison shows that alcohol have more effect on having higher levels of cancer.

The last comparison that will be done is between number of patients with heavy smokers and heavy alcohol drinkers. 43.5% of the patients are heavy alcohol drinkers and %35.6 of the patients are heavy alcohol drinkers. We can state that smoking have more effect on having the cancer whereas alcohol causes higher levels of cancer.

Acknowledges

References

-"Lung cancer data" by prithivraj: https://data.world/cancerdatahp/lung-cancer-data/workspace/file?filename=cancer+patient+data+sets.xlsx

-"Environmental and occupational causes of cancer: A call to act on what we know" by Richard W. Clapp, Genevieve K. Howe, Molly M. Jacobs: https://www.sciencedirect.com/science/article/abs/pii/S0753332207001710

-"7 thing to know about cancer" by Danielle Underferth: https://www.mdanderson.org/cancerwise/does-alcohol-cause-cancer.h00-159383523.html#:~:text=The%20ethanol%20in%20alcoholic%20drinks,levels%20of%20hormones%20like%20estrogen.

-BBC's article "Non-communicable diseases": https://www.bbc.co.uk/bitesize/guides/z372ng8/revision/11