The Impact of Climate Change on Agriculture: Modeling Wheat Production and Projecting Insurance Trends in Selected Counties in Southwestern Kansas and Northeastern Washington

2019-20 Modeling the Future Challenge Project Report
February 28, 2020

Executive Summary

Agriculture is dependent on a wide range of ecosystem processes that support productivity including maintenance of soil quality and regulation of water quality and quantity. Multiple stressors, including climate change, increasingly compromise the ability of ecosystems to provide these services. Key near-term climate change effects on agricultural soil and water resources include the potential for increased soil erosion through extreme precipitation events, as well as regional and seasonal changes in the availability of water resources for both rain-fed and irrigated agriculture.

The predicted higher incidences and greater severity of extreme weather events will have an increasing influence on agricultural productivity. Extremes are influential because agricultural productivity is driven largely by environmental conditions during critical threshold periods of crop and livestock development. Improved assessment of climate change effects on agricultural productivity requires greater integration of extreme events into crop and economic models.

Wheat is one of the world's most significant commodities; it is in a large portion of the foods we eat daily, but it has begun to suffer from the negative effects of climate change. The impacts are severe: wheat crops are dying and production levels are falling. Therefore, it is of importance to investigate these effects in order to determine the most important causes of loss, predict future trends of extreme weather events, and give recommendations to insurance companies, public policymakers, and local farmers in the states of Kansas and Washington in order to help them successfully combat them. This project focuses on a region in Kansas, but applies models developed for Kansas to counties in Washington to observe the performance differences between disparate locations. Both primary and optional data sources were utilized to analyze cause of loss, extreme weather events, temperature patterns, and more.

First, to observe the distribution and frequency of causes of loss, multiple box-and-whisker plots were created in addition to histograms and pie charts. From these, it was evident that drought was the most common cause of loss in the dataset, and thus, water scarcity and rising temperatures were key factors that characterized climate change in these areas.

In order to aid insurance companies, loss ratio was also predicted, and based on the project trend, recommendations were provided to curb projected issues. Factors that were necessary for calculating loss ratio included indemnity, liability, and earn premium rate, which were included in the datasets. Wheat production was also predicted, by taking climate changes over time into account through multi-linear regression models.

In summary, similar modeling techniques involving multiple linear and polynomial regressions were applied to geographically distant regions with some shared features, and were tweaked to ensure good performance despite different underlying regional feature topologies.

Background Information

In addition to being a key resource for human livelihoods, agriculture is a major source of wealth for the United States economy, bringing in about $300 billion a year in commodities, with livestock accounting for approximately half of that value. Production, yield, and development of these commodities is vulnerable to changes in climate through direct factors such as variation in temperature or precipitation levels, or indirect factors such as pests or availability and success of ecosystem services (a combination of provisioning, reguating, supporting, and cultural services that contribute to human well-being).[5] Therefore, agriculture relates productivity, climate change, and ecosystem services as a web of complex interactions between different factors.

Because of agricultural production's sensitivity to changes in weather conditions, climate change presents difficult challenges to U.S. farmers, agricultural workers, and the economy, as selling prices are variable to the production of the commodity. To adapt to changing circumstances, we must act quickly in order to maximize production while also minimizing costs associated with negative effects of climate change. There are a variety of ways to address this overarching problem, from adjusting planting patterns and soil and water management to current weather conditions, to altering international trade and involvement of federal programs. Responding to the threats that climate change brings requires aggregating diverse factors such as production and consumption levels from the public, education and research, insurance services, and government policies. Understanding how to combine these factors to create effective responses to climate change is imperative.

The way a plant reacts to climate change is determined by several components, including carbon dioxide and oxygen levels, precipitation indices, solar radiation, and atmospheric temperature, all of which are impacted by changing climate.[5] All crops have different thresholds of these factors that dictate their success (the capacity to develop correctly, grow, and reproduce), and this is why plants are currently grown in areas where their ideal temperature needs are met. Current temperature predictions reveal that by 2100, increases will reach 2.8 to 3.2 degrees Celsius even with the help of control policies.[6] As temperatures and carbon emissions from greenhouse gases and fossil fuels continue to increase, the locations at which crops are currently planted will shift to accommodate optimal temperature ranges.

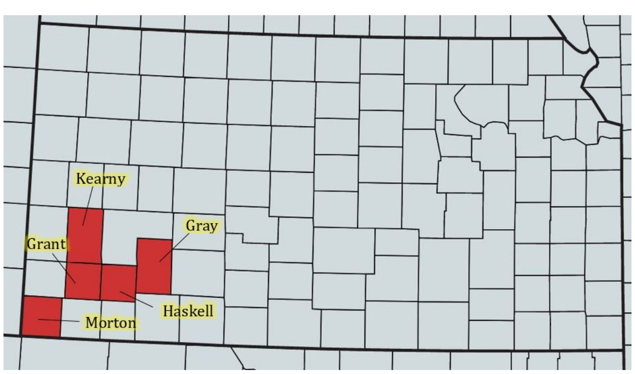
Wheat, the commodity studied in this project, is arguably the most important staple crop in the world, accounting for a fifth of globally consumed calories, and it continues to rise in demand. In addition to being a major source of starch and energy, wheat also provides substantial amounts of many components which are essential for health, notably protein, vitamins, and dietary fiber.[1] It is also the basis of many common foods such as breads, cereals, crackers, cakes, pizza, and pasta. Here, the relationship between wheat production and climate change in select counties of both Kansas and Washington (leading states in production of wheat) was investigated to discover possible methods of combating negative effects of climate change on this important crop.

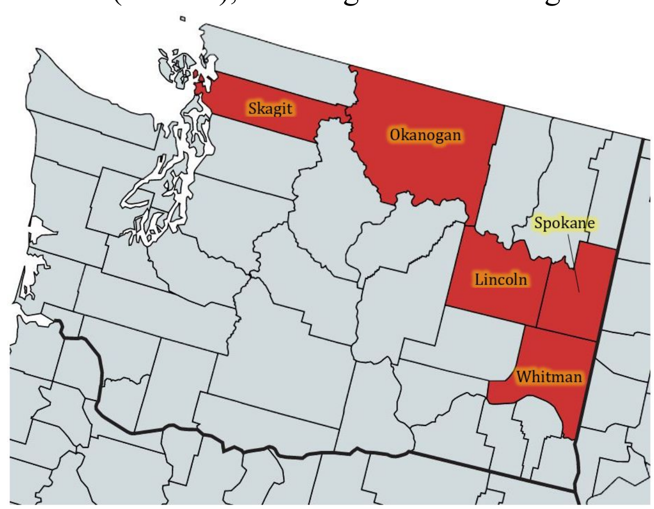
In this project, five southwestern counties in Kansas (Kearny, Grant, Morton, Haskell, Gray) and five northeastern counties in Washington (Whitman, Okanogan, Skagit, Spokane, Lincoln) were studied, in part to compare and contrast modeling techniques for forecasting wheat production and understanding trends related to profits and losses of insurance companies in this space. Reasons for this methodology are further discussed below.

The state of Kansas leads the United States in wheat production, and all counties in Kansas grow it. Sometimes called “the wheat state,” around one-third of Kansas’ 60,000 farmers are wheat farmers. Kansas’ economy relies heavily on wheat, with about one-fifth of the country's total wheat production coming from it. Three of the five varieties of wheat grown in the US are produced in Kansas: hard red winter (e.g. bread and rolls), soft red winter (e.g. flatbreads, crackers, and pastries), and hard white (e.g. noodles and pasta).[4]

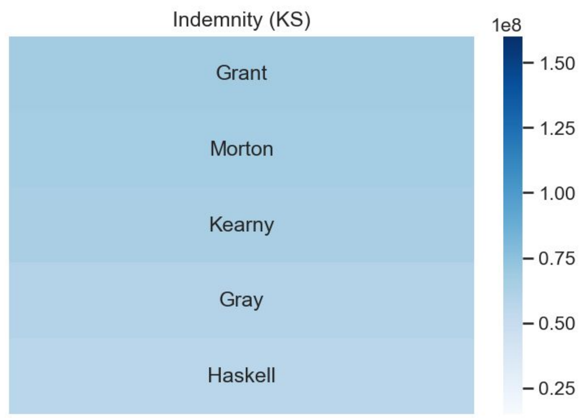
Kansas is known as the Wheat State, or the “breadbasket of the world” for good reason. According to statistics in 2017, 7.6 million acres of wheat were planted, and 6.95 million acres were harvested, making an average yield of 48 bushels per acre.[7] Even with wheat’s great success, it is not Kansas’ only main crop; in fact, wheat only accounts for 8.6 percent of the state’s total agricultural production. However, approximately 19.2 percent of the nation’s total wheat comes solely from Kansas — almost a fifth![1] Environmental factors including climate, soil, and rainfall, in addition to central geographical location in the United States help Kansas to maintain its role as the nation’s leader in wheat production. According to the Kansas Department, the wheat industry in Kansas outputted $1.44 billion and generated 3,215 jobs throughout the state in 2019. In addition, the agricultural industry as a whole supports a total of 10,487 jobs and creates a total economic contribution of approximately $2.57 billion.[7] Currently, Kansas is likely the world’s best most abundant source for hard red winter wheat.

While attempting to understand how and why wheat production varies over time, it becomes meaningful to test modeling techniques on various regions with significant wheat production. Another state, quite geographically distant from Kansas, is also a significant producer of wheat — Washington. The state of Washington mainly produces soft white and club varieties of wheat (e.g. pastries, crackers, cookies). Additional varieties include hard white (e.g. blended flours and certain types of noodles) and hard red winter wheat.[4] Eastern Washington is home to some of the best wheat quality grown in the world. In 2018, Washington wheat growers harvested 2.2 million acres of wheat which had an average yield of 70.8 bushels per acre. Total wheat production for 2018 was 153.2 million bushels. Washington is ranked fourth in the nation’s top wheat producing states. In addition, wheat ranks third in the Washington commodities based on production value, representing nearly $691 million.[8] Washington is also one of the nation’s leading wheat-exporting states, with 85 to 90 percent of its production exported each year. Compared to the national statistic of a 46% exportation rate (in 2018), Washington is a thriving force in wheat production.[8] Below are maps indicating the counties selected for this study, in southwestern Kansas and northeastern Washington.





A main difference between the two regions above was their counties' past relative indemnities. The heatmaps below depict the payment indemnities in U.S. dollars of selected counties in Kansas: Grant, Morton, Kearny, Gray, and Haskell. As evident by the nearly-identical shades of blue between the five counties, Kansas has extremely little variation in its payment indemnities. All counties are centered in the $50,000,000 to $75,000,000 range, and they all produce high quantities of wheat.



The default model construction, a composition of regressions, would likely work well in projecting trends for counties with similar characteristics of the selected Kansas ones, but a greater goal was to design and develop the model in a way such that it could be applied to locations across the nation, including those with low production and those with even higher wheat production as well. On the right, there is a heatmap of payment indemnities of selected counties in Washington: Whitman, Lincoln, Spokane, Okanogan, and Skagit. This plot differs greatly from its Kansas counterpart. There is a much larger gradient in the shades of blue that represent indemnities; Skagit and Okanogan have low indemnities (and thus, little wheat production), as they are both under $25,000,000. Spokane has close to $50,000,000. Finally, the two highest counties, Lincoln and Whitman, have indemnities over $1,250,000,000. The motivation for choosing these counties arose from the fact that Washington had such great variation in its wheat production indemnities, and by inputting more diverse data into the model, the model would be able to make more knowledgeable, and therefore better and more applicable, regional-level predictions.

The goal of this project was to demonstrate that predictive models could work well for states with varying levels of wheat production. If a model were trained on only data from Kansas, a wheat-rich state that grows lots of wheat throughout all counties, it would likely only effectively predict production yield in states with similar conditions. In other words, the model would likely be accurate and usable only for areas with similar wheat production to Kansas. Therefore, counties from a second state were incorporated: Washington. Washington is also a large producer of wheat, but there is much more variance with respect to wheat production between its counties. Northeast Washington has a "gradient of counties," in terms of wheat production, meaning that a few counties grow tens of millions of bushels, whereas others grow much less. Kansas and Washington are also located in different regions of the US; Kansas is in the central US while

Washington lies in the Northwest. By training models on two vastly different datasets, in terms of the amount of wheat produced and geographic location, it is possible to predict crop yield and production more accurately given an arbitrary input.

Wheat makes an important contribution to the human diet and is one of the most commonly consumed cereal grains in the world; it is a staple ingredient of many baked goods, such as bread. Unfortunately, it is beginning to suffer from hotter temperatures and water shortages. According to international scientists, if climate change is not mitigated, 60% of wheat-growing areas will have water scarcity by 2100, which is a major increase from the current 15%.[2] A study at Kansas State University also revealed that for every degree Celsius increase in temperature, wheat production would be cut by 6-8%, resulting in millions of tons of lost wheat. Although the UN hopes to limit climate change to an increase of 1.5 degrees Celsius, this requires dramatic, and possibly unrealistic, action.[3]

Climate change in both Kansas and Washington could be extremely deadly for wheat crops as well as other crops. Although the wheat industries in both these areas have experienced great successes, they continue to face challenges that may prevent more growth. In Kansas, recent studies conducted by the federal government have targeted the agricultural sectors of Midwestern states (including Kansas). This national study, called the Fourth National Climate Assessment, forecasts that rising temperatures and unpredictable weather patterns like flooding and droughts will continue to increase.[9] It explains that any changes in climate will quickly pose major challenges to agricultural production and growth, because the crops will suffer from failure and altered rates of external pressure from pests, weeds, and disease. Rural communities and farms in Kansas are very interlocked with the agricultural sectors, so they are the most threatened and will likely face the most damage.

For the past century, most of Kansas has been increasing in temperature by at least half a degree Fahrenheit per year.[9] The soil is becoming drier, and rainstorms, which lead to floods, are becoming more frequent and intense. Warmer winters and changes in the timing and length of rainfall events have negatively impacted crop yields, especially wheat yields. As the years progress, summer months are going to be hotter and drier, creating huge obstacles for successful agriculture.

Ironically, changing climate will increase temperatures resulting in an increased demand for water, but the availability of water will decrease simultaneously. Soil will continue to dry out as it has been doing over the last several decades, because of more evaporation, plants' intake of water, and the fact that rainfall in the summer will go down. As projected by Kansan studies, seventy years from now, long periods without rain could last three to four days longer than they do now.[10] Warmer temperatures cause the average flow of water in rivers and streams to decrease as well; therefore farmers will have less-than-enough water available to irrigate their crops. About 22% of farmland in Kansas is powered by irrigation, which is mainly groundwater, and because of climate change, levels of groundwater are decreasing. In fact, the amount of water has already decreased to 75% of its original quantity in the 1950s in many locations in Kansas.[10] At the rate of current climate change, farmers will soon suffer drastically due to the inadequate water source. Water shortages in combination with higher temperatures will cause production yield rates of wheat in Kansas to decrease dramatically.

In Washington, changing climate conditions will alter the geographic areas in which specific crops can be grown. Therefore, wheat could be relocated to accommodate warmer climates. Crop productivity will be affected by several factors, including changes in average temperature, extreme natural events, elevated carbon dioxide levels, availability of water, and stress from weeds, pests, and invasive species. Climate change in Eastern Washington will not be very severe, meaning that impacts to crops will be minimal at least one to two decades from now.[11] There will be increased concentrations of carbon dioxide, and this will compensate for some of the negatives of climate change, as plants will then be able to produce more glucose and oxygen. However, this is no reason to ignore or downplay the effects of climate change, because climate impacts will continue to worsen over time.

Wheat is sensitive to rising temperatures and changes in water availability. Longer growing seasons, warmer temperatures, and higher carbon dioxide concentrations may increase productivity for some crops. In a recent Washington study, winter wheat productions were taken at different elevations, both with and without irrigation, and the best yields were in areas with a lot of rainfall, temperate conditions, and at elevations from 1000 to 1500 meters.[11] Both non-irrigated and irrigated harvests have increased with global warming, which has also allowed for increased production at higher elevations. The harvests also improved with the presence of higher levels of carbon dioxide.

Different crop zones across Washington support different commodities and agricultural practices, and these zones are likely to have different responses to climate changes; regardless, they will probably be negatively affected in some way by limited water, more weeds, pests, diseases, heat, drought, and flooding. Changes in climate may affect which crops can grow efficiently in the state, and they could soon shift from their current region to now-cooler areas in the near future. The crops that are most threatened by climate change in Washington are those that have a small range of optimum temperature and those that cannot adapt to new conditions quickly and effectively.

Data Methodology

Data from four different sources was utilized in this project, including the primary data (Cause of Loss, Report Generator) and two supplementary sources (NOAA Historic Climate Data, USDA National Agricultural Production Information). These supporting datasets were chosen because the goal was to observe the effects of climate change on production, yield, and sales of wheat, as well as to determine the environmental and climate factors that were most influential to wheat in order to make predictions and recommendations for insurance companies. The model's final prediction task was to estimate the loss ratio for an insurance, which is calculated by dividing indemnity by total premium. By using historic climate data in conjunction with agricultural production and insurance policy information, this was possible.

The Cause of Loss dataset was provided by the USDA Risk Management Agency and provided information on causes of loss for wheat crops, liability of the insurance company, subsidy, indemnity, and more. This data was monthly from 1989 to 2019, and it recorded all insurance policies that had a loss. However, it did not record policies without a loss. 10 datasets were retrieved: one for each of the five counties in both Kansas and Washington. This data is critical for the project because it categorizes insurance losses and crop failures based on a certain cause of loss; therefore, it is possible to perform analyses to determine which climate factors are most impactful to wheat and will help make predictions about which factors to target and combat. This data is organized into several columns:

- State (state ID; 20 for Kansas, 53 for Washington)
- State name (KS for Kansa, WA for Washington)
- County ID (specific to county)
- County name (name of the county)
- Year (from 1989 to 2019)
- Month (number from 1-12 representing each of the twelve months of the year)
- Commodity (wheat)
- Cause of loss (variety of options; significant causes include drought, wind/excess wind, cold winter, freeze, cold wet weather, plant disease, failure of irrigation supply, excess moisture/precipitation/rain, hail, hot wind, frost, heat, insects)
- Payment indemnity in US$ (the total amount of the loss for the designated loss event)
- Payment acreage (number of acres damages by the cause of loss)
- Liability in US$ (the maximum dollar amount of insurance for the crop which the insurance is required to cover)
- Subsidy in US$ (amount of money granted by the state/government to assist payment for the damaged crops)

The Agricultural Report Generator, also from USDA, is a summary of business that allows the users to isolate a specific commodity (crop) and location down to the county level to analyze many different factors of that commodity. The data can then be downloaded as an Excel spreadsheet. The major difference from the previous dataset is that it includes information on all crop insurance policies, not just those with a loss. However, the limitation to this dataset was that it was recorded on an annual, not monthly scale.

The valuable information gained from the Report Generator discusses many more money-based measurements, and gives information on insurance premiums, which is the amount of money an individual or business pays for an insurance policy. These insurance premiums are necessary for computing loss ratio, which will help make recommendations for insurances and businesses. Ideally, insurances want their loss ratios to be less than 1; otherwise, if the loss ratio is greater than 1, the insurance is losing money, as it is paying more for the policy than it is receiving in premium. Loss ratios are usually higher when liabilities and indemnities are both large. The columns utilized in this dataset are:

- Commodity year (year from 1989 to 2019)
- Commodity name (wheat)
- State abbrv (KS for Kansas, WA for Washington)
- County name (name of county)
- Policies sold (total number of crop policies sold; units that had a loss)
- Policies earning prem (number of crop policies insured with a premium)
- Policies indemnified (number of crop policies with a reported loss)
- Units earning prem (total number of units insured with a premium)
- Units indemnified (total number of units indemnified regardless of whether the unit was insured with a premium or not)
- Quantity (number of acres insured that the insurance company is liable for)
- Quantity type (acres)
- Liabilities in US$ (the maximum dollar amount of insurance for the crop which the insurance is required to cover)
- Total prem in US$ (total premium amounts the insurance company is receiving from policies with and without a loss)
- Subsidy in US$ (amount of money granted by the state/government to assist payment for the damaged crops)
- Indemnity in US$ (the total amount of the loss for the designated loss event)
- Earn prem rate (amount of premium earned divided by liability)
- Loss ratio (ratio of losses to premiums earned)

The third type of data used was monthly climate and temperature data from 1989 to 2019, gathered from the National Oceanic and Atmospheric Administration (NOAA). From each of the ten counties studied, four different measurements were collected: average temperature, minimum temperature, maximum temperature, and precipitation values. In addition, two statewide measurements were retrieved, which were the Palmer Drought Severity Index (PDSI) and the Palmer Hydrological Drought Index (PDHI). PDSI indicates the dryness of an area, where -10 is extremely dry and 10 is extremely wet. Similarly, PDHI measures the hydrological impacts of drought on things like reservoir levels, groundwater levels, etc. By analyzing climate data and correlating it with the other types of data discussed, it is possible to see how climate change affects temperatures over time and whether they are getting warmer, as predicted by climate change. There are three columns in all datasets, and they are as follows:

- Date (year from 1989 to 2019 followed directly by month from 1 to 12)
- Value (drought index on a scale of -10 to 10)
- Anomaly (the distance from the value recorded the mean value)

The final source of data was from USDA's Agricultural Production Information in the Quick Stats data portal. This data was recorded by year, but only certain years were available for some statistics. Also, some measurements allowed for data to be taken at the county level, whereas others were only available at the state level.

When at the portal, six types of data were collected. The first two were the number of acres harvested and the number of acres planted. Using these measurements, it was possible to see the success rate of wheat crops and what percentage was grown successfully for consumption and selling. Using the success rate, it was also straightforward to compute the loss rate, which can be defined as the portion of wheat that was damaged and not able to be grown successfully. Both of these statistics were taken at the county level.

The next two measurements relate to wheat production: production measured in dollars and production measured in bushels. The former was only available at the state level, but the latter was available at the county level. These measurements were investigated because it is important to be able to analyze how much wheat was produced compared to how much was sold and consumed. Then, data for yield of wheat was retrieved. This measures the yield of wheat in bushels per acre, which is a convenient method to determine the successes of individual areas and farms. This would help target intervention to specific places even better.

Lastly, measurements for sales of wheat in dollars were analyzed. This statistic is somewhat similar to the production dollar measurement, but it differs in the fact that it accounts only for crops that were sold to the public, not the total of all that were produced. Therefore, this measurement will always be less than its corresponding production value. All the datasets explained above follow a specific column format:

- Program (Census Bureau survey)
- Year (year from 1989 to 2019; will vary in each dataset)
- Period (year)
- Geo Level (either state or county depending on the dataset)
- State (Kansas or Washington)
- State ANSI (state ID)
- County (name of the county)
- County ANSI (county ID)
- Commodity (wheat)
- Data Item (the corresponding measurement name based on the several described above)
- Value (the corresponding value of the measurement based on the several described above)

Mathematics Methodology

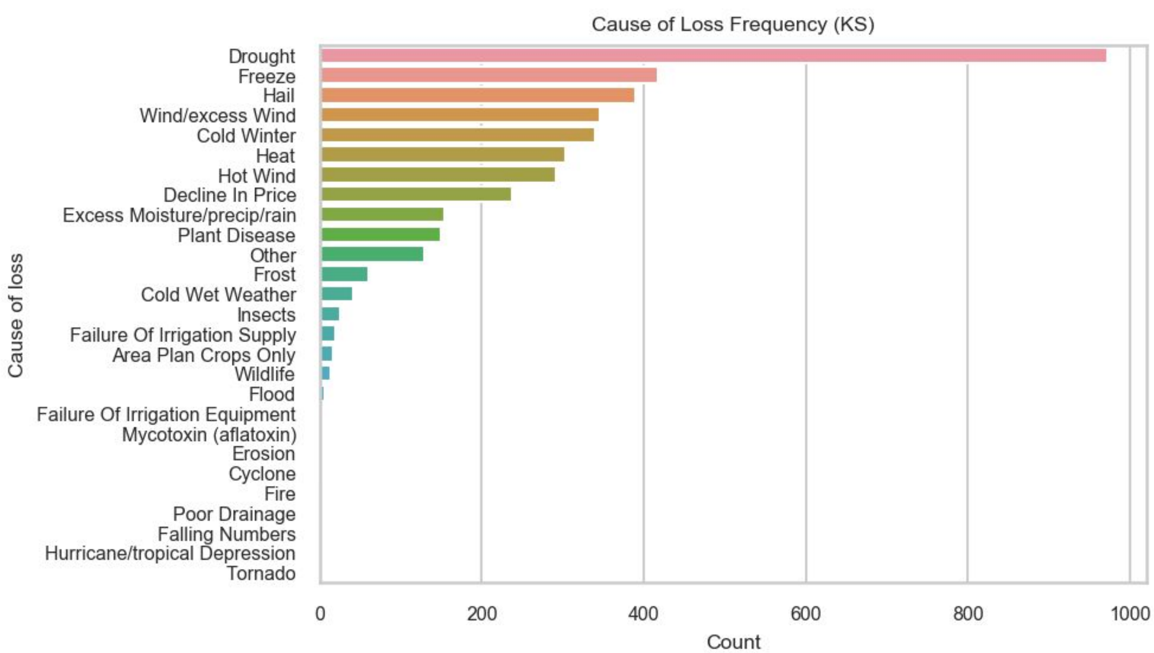
A goal of this study is to construct a model to predict insurance loss ratios based on changing factors of climate change and historic insurance data and use those predictions to suggest recommendations for insurance companies and policies. There are several key measurements that are necessary for calculating loss ratio; they include indemnity, which is the sum of money an individual or business receives for qualifying losses paid under an insurance policy. Indemnity will compensate for the rest of the money not covered by a deductible or government subsidy, up until the amount that the insurance guarantees. The second measurement needed is premium values. Premiums are the amounts of money an individual or business pays for an insurance policy for risk protection. Loss ratios are used to measure the success and profitability of a company. With these, loss ratio is calculated as shown below:

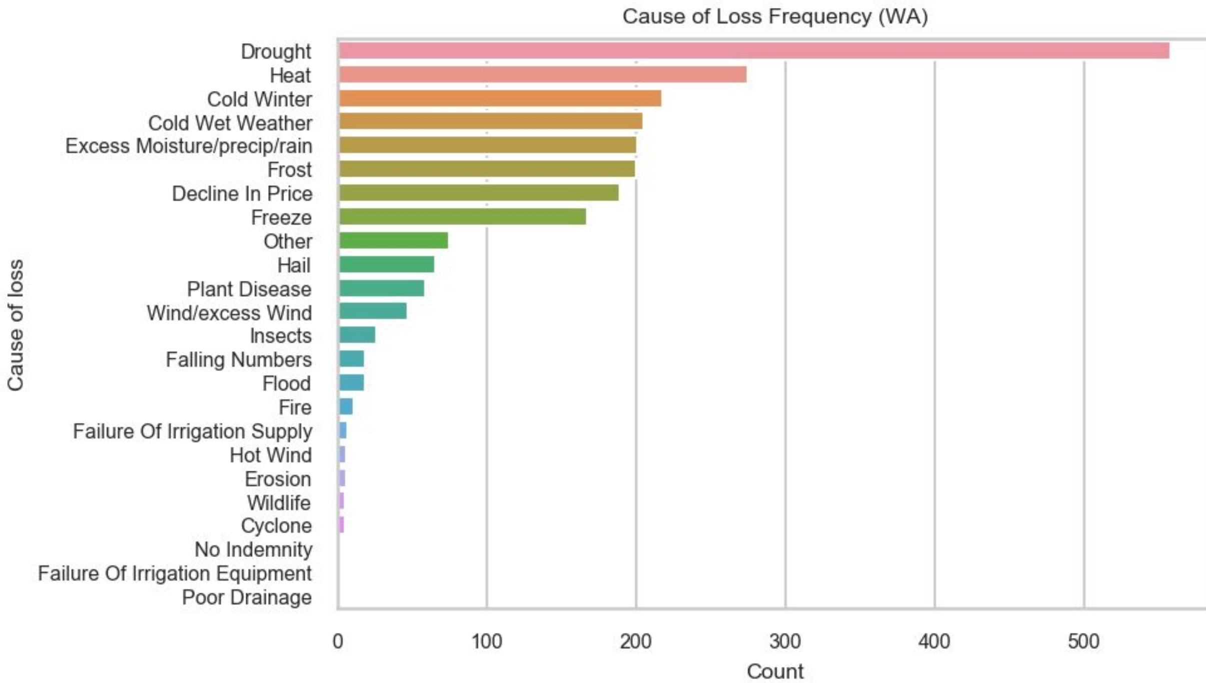
$$\textit{loss ratio} = \textit{indemnity}\ (\$) \div \textit{total premium}\ (\$)$$

To dive deeper, we must note that total premium is the product of two factors: liability and earn premium rate. Liability is the maximum dollar amount of insurance for the crop which the insurance is required to cover. It may also be defined as the dollar amount the the insurance company pays to the farmer when there is zero yield. The earned premium rate is the amount that an individual or business pays to an insurance company to cover for the insurance policy. Below is the calculation for total premium:

$$\textit{total premium}\ (\$) = \textit{liability} \times \textit{earn premium rate}\ (\$)$$

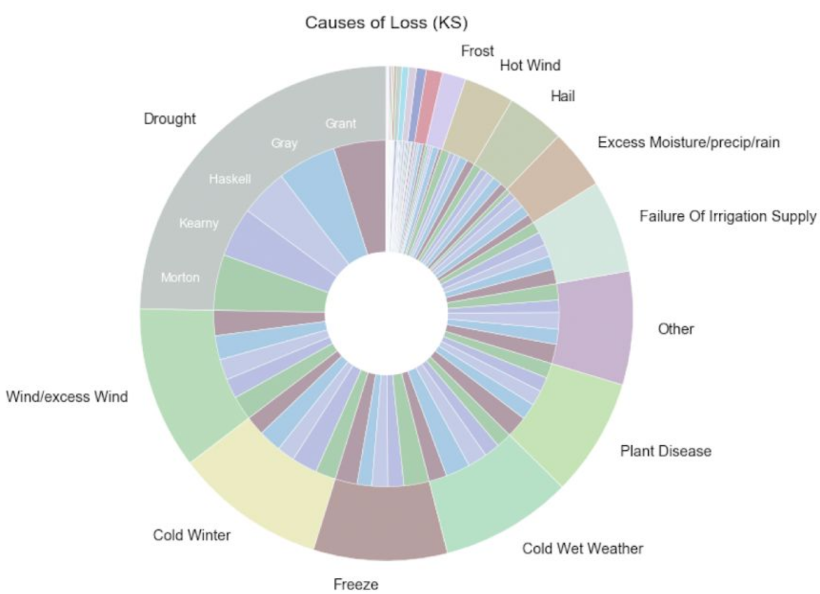
By analyzing the distribution of causes of loss, it is possible to incorporate the specific causes of loss that contribute most to wheat damage, to eventually predict future losses:

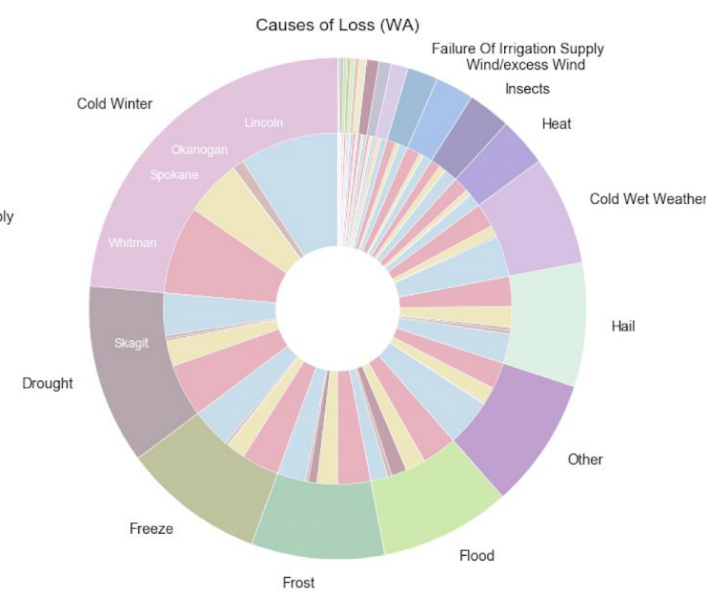




We see from the histograms above, depicting cause of loss frequencies for wheat production from 1989-2019, that the most common causes of loss among the selected counties in Kansas were drought, freeze, and hail, and for the selected counties in Washington, they were drought, heat, and cold winter. These findings were incorporated in the predictive models for (1) indemnity, (2) earn premium rate, and (3) liabilities, which were then used to calculate loss ratios. In particular, each county in each year was associated with a vector containing that county's contribution (in percentage) to the top three causes of loss for that region (either southwest Kansas or northeast Washington).

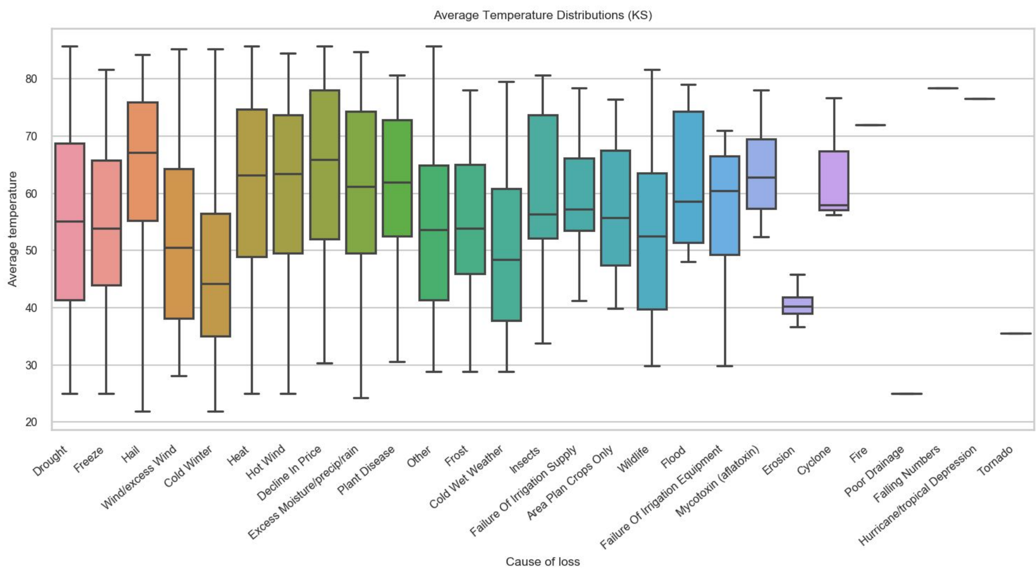
For example, these nested pie charts reveal the relative compositions of causes of loss at the county level. We see that all the counties in the selected region of Kansas contribute fairly equally to each cause of loss, perhaps indicating that predictions made about how climate affects

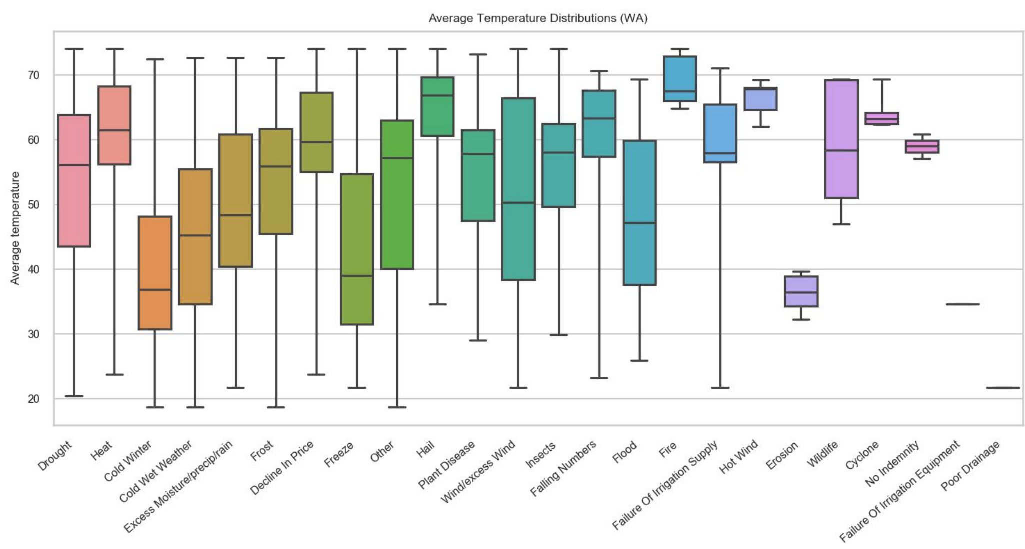




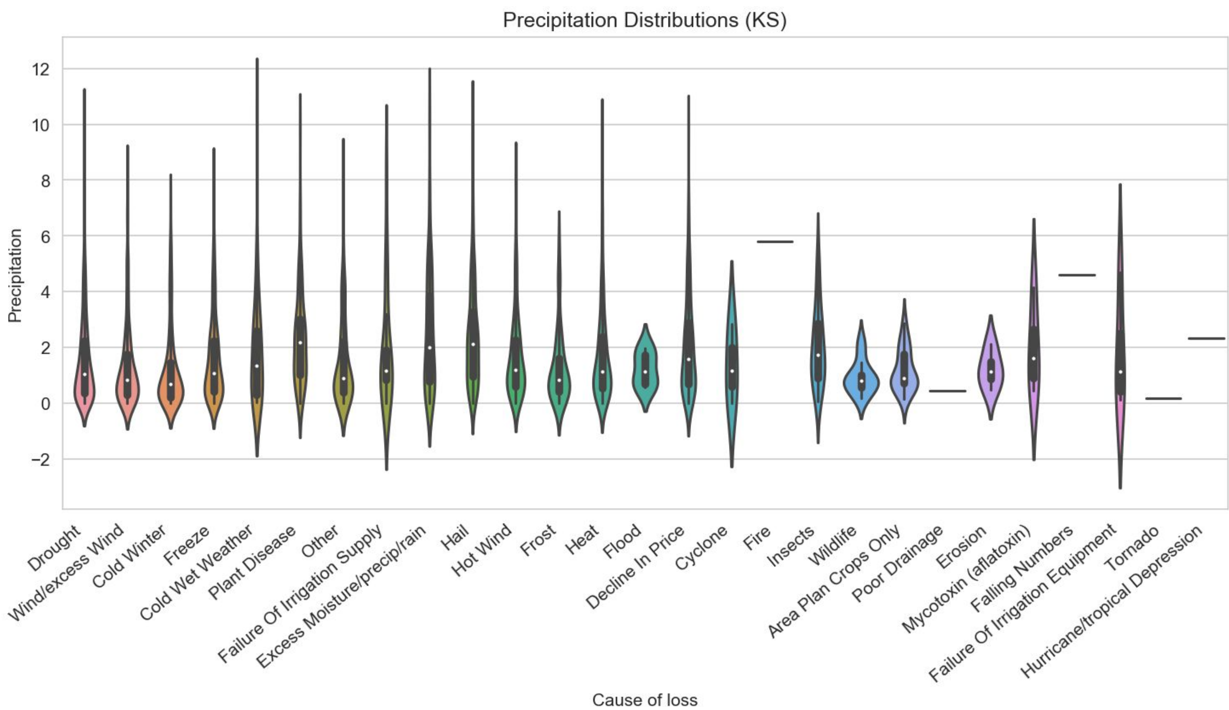
wheat production at the regional level would be similar to those made at the county level. (Constrastingly, this is not the case for the Washington region at all, as colored sector size ratios differ greatly by causes of loss.)

It appears that drought is a major factor for which recommendations are needed. To discover other major factors that are also beneficial for future trend prediction as well as the distributions of causes of loss when plotted against quantities such as temperature, drought levels, and precipitation levels. Understanding these relationships will allow the final model, built to predict insurance loss ratio, to account for various types of climate factors as well and accurately as possible. To achieve these visualizations, box-and-whisker and violin plots were created. Fundamentally, they can be explained as a hybrid of a box plot and a kernel density plot, and they show distributions and probabilities in the dataset.

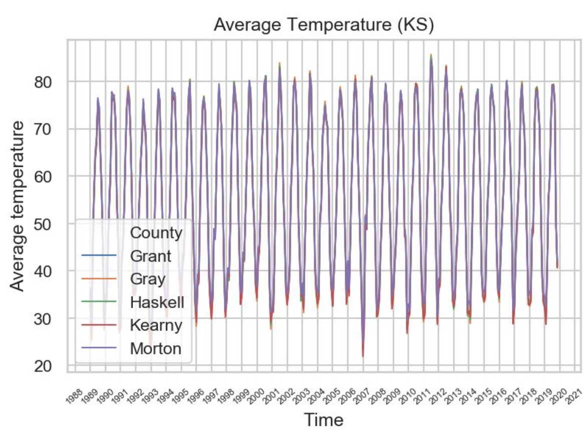


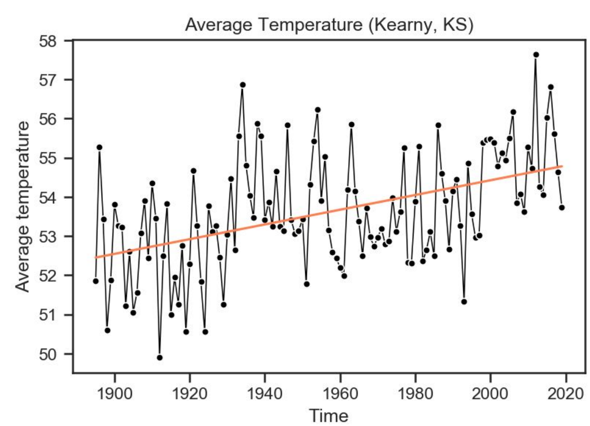


The box-and-whisker plots above for the selected Kansas and Washington regions show the relationship between causes of loss and associated average temperatures in Fahrenheit.



The violin plot above of the Kansas region shows the relationship between causes of loss and Palmer Hydrological Drought Index (PHDI). This drought index, on a scale of -10 (dry) to +10 (wet), uses precipitation and temperature measurements to estimate relative dryness of an area.





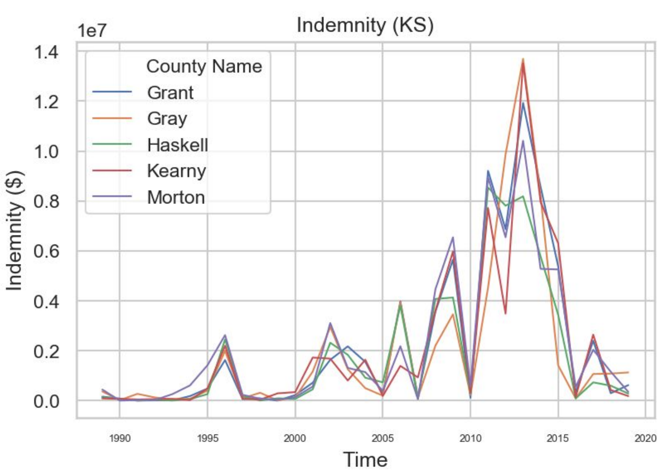
When looking at temperature data from a monthly scale as displayed in the graph on the left, there seemed to be little relationship between time time and temperature. It seemed to stay constant over time, peaking in the summer months and dropping in the winter months as expected. However, when analyzing it on an annual scale, it is much easier to see its progression over time. Using a wider lens to analyze temperature reduces noise that is produced at the granular level. There is a clear upward trend in the dataset, indicating that temperature is increasing likely due to global warming and climate change.

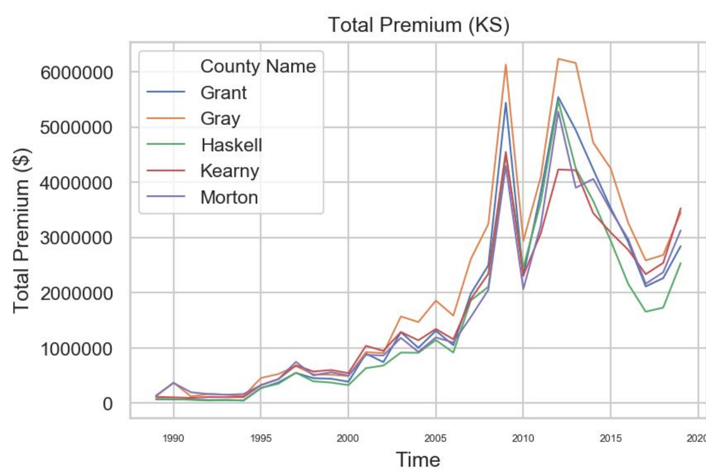
To evaluate how much climate has impacted temperature, I found the change in temperature from 1989 to 2019, from the line of best fit equation, where y represents average temperature in degrees Fahrenheit and x represents the numerical year:

$$y = 0.019x + 16.835$$
$$0.019(1989) + 19.835 = 54.219$$
$$0.019(2019) + 19.835 = 54.782$$
$$0.019(2050) + 19.835 = 55.365$$

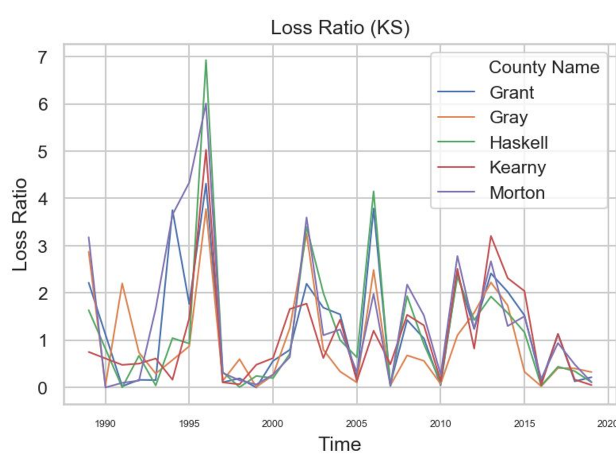
At this rate, by 2050, the average temperature will reach 55.365 degrees Fahrenheit. Although this may seem like a small shift, wheat is a sensitive crop, and production reduces by 4.1% to 6.4%, leading to a loss of millions of bushels. Therefore, temperature needs to be monitored to prevent further losses to wheat production.

We observe the trends of the targets to predict — indemnity and total premium (the product of liabilities and earn premium rate) — as well as loss ratio itself. In this indemnity graph of the Kansas region, indemnity values seem to stay quite constant across counties until around 2009 and 2010, when they begin to fluctuate greatly. This implies that there must have been a trigger or cause for this change. However, values drop quickly back down to near their original level a few years later in 2015 and 2016. Therefore, it is best to note that the four years within the fluctuation period may be classified as outliers. Meanwhile, total premium is largely on the rise over time:

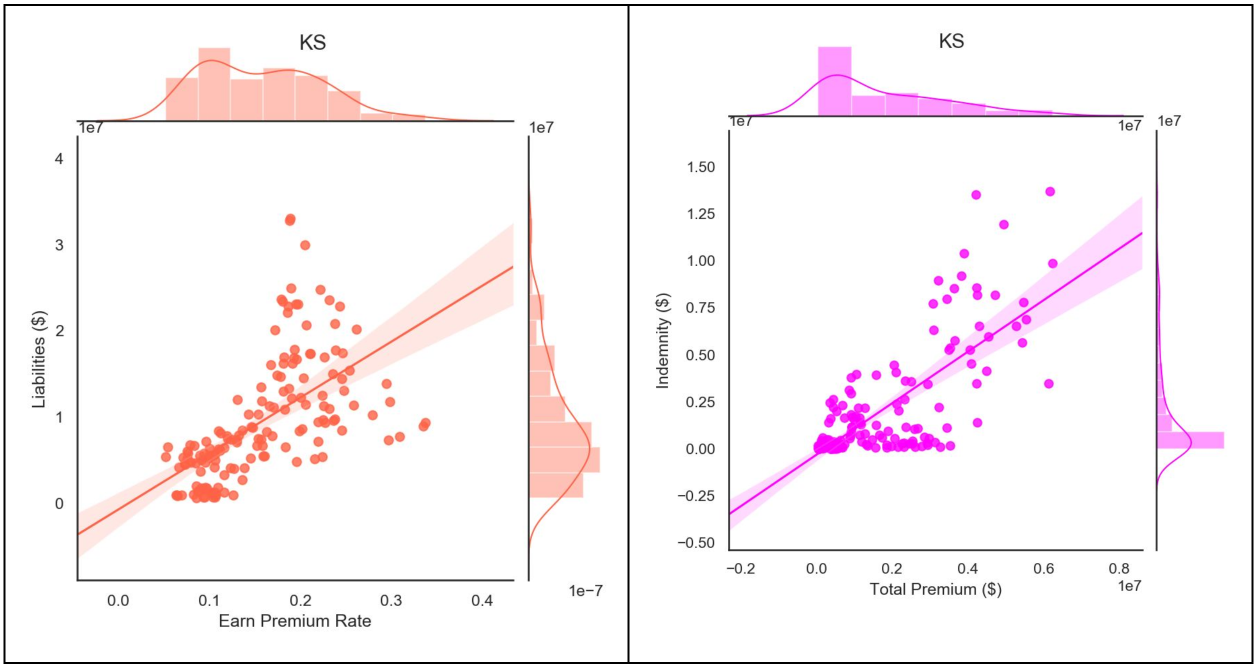




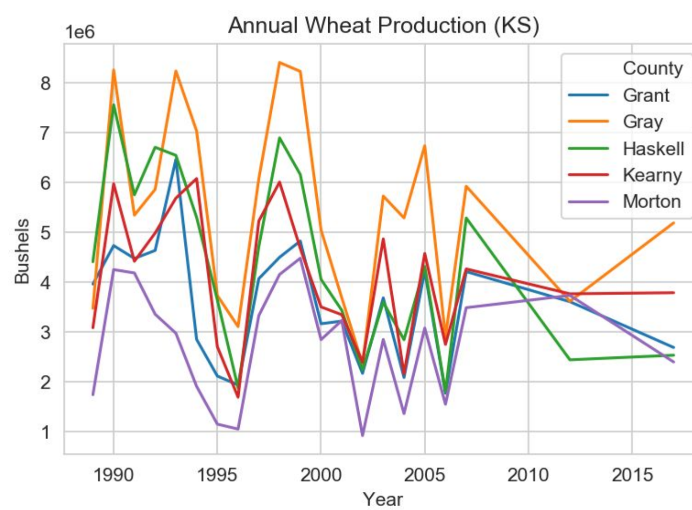
At the right is the plot of loss ratios (all selected counties in KS), which are elements we'd like to predict into the future, to advise insurance company policy. As is evident from the graph, the trend is very difficult to estimate from simple regressions, so a multiple regression was used to resolve this.



A confirmation that total premium is increasing quickly over time is by this liabilities versus earn premium rate jointplot. For total premium to be constant, either liability or the earn premium rate would have to move in the opposite direction as the other one (if the earn premium rate, increases, liability would have to decrease). This is not the case; the product of the liability and the earn premium rate is generally increasing. Moreover, in the indemnity versus total premium jointplot to the right, both variable distributes are similarly left-skewed. We also see that for several total premium values, the indemnity remains unchanged, indicating that their ratio (the loss ratio) fluctuates, sometimes wildly, sometimes not.

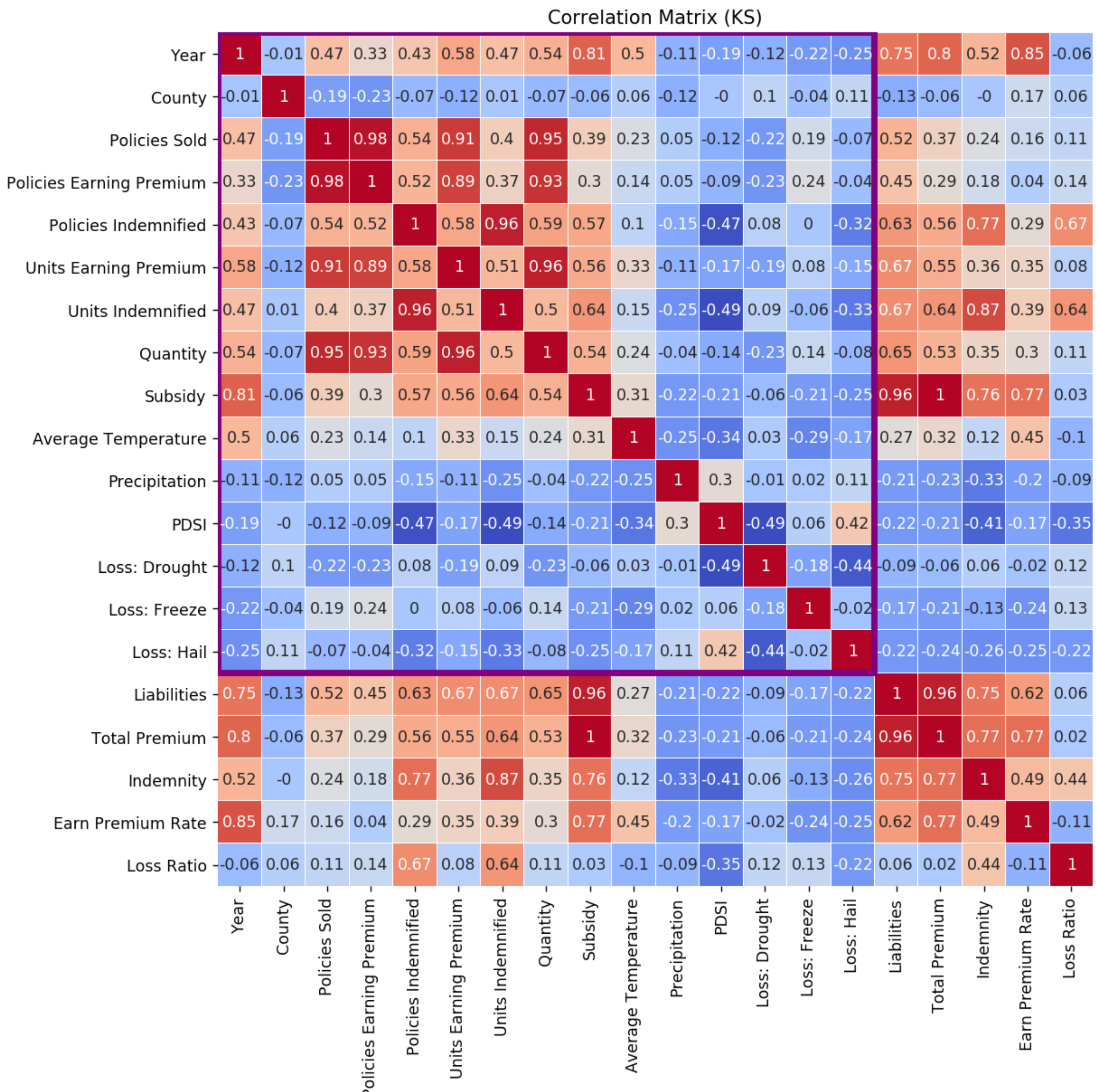


This graph shows the relationship between time and number of bushels produced in Kansas from 1989 to 2019, in year-long intervals. It's incredible how well the counties Grant, Gray, Haskell, Kearny, and Morton are in sync with one another. The fact that they all line up with each other implies that there are common climate factors (causes of loss) that affect multiple counties in the same region. Therefore, the results and recommendations that we draw from one county can easily be applied to a different country within the same region of Kansas. (Contrastingly, this is not the case for the Washington region at all, where wheat production curves differ wildly between counties. We hope to design our model cleverly so as to give accurate, robust predictions.)



Model Results

To predict trends through 2050 in order to provide recommendations to insurance companies for how to best gear their policies, regression was performed with all twenty features, for a total of 400 R^2 values. The goal of this process was to find the most important features, which would then be put into regression models to predict liability, earn premium rate, and indemnity. These were the chosen target factors of interest, because they are necessary for computing loss ratio and total premium, as shown previously.



Each cell in the heatmap above represents the individual correlation coefficient between two features. Values located outside the purple outline are measures we are attempting to predict (the targets), and values inside the outline are the features inputted into the regression model. It is important to focus on the correlation coefficients inside the purple box that are high because these designate a strong, predictable correspondence between two factors. (To avoid multicollinearity problems, we choose a single feature from pairs that have strong correlations.) By analyzing which factors have the greatest proportionality with others, we will be able to target certain ones and take action appropriately; we will be able to narrow our strategies to combat climate change. Using the information that we gain from correlations inside the purple box, we can then apply that to predict values such as total premium, indemnity, and ultimately, loss ratio. Analyzing these relationships will be helpful in making recommendations to insurance companies later in this report. Some examples of strong correlations are listed below:

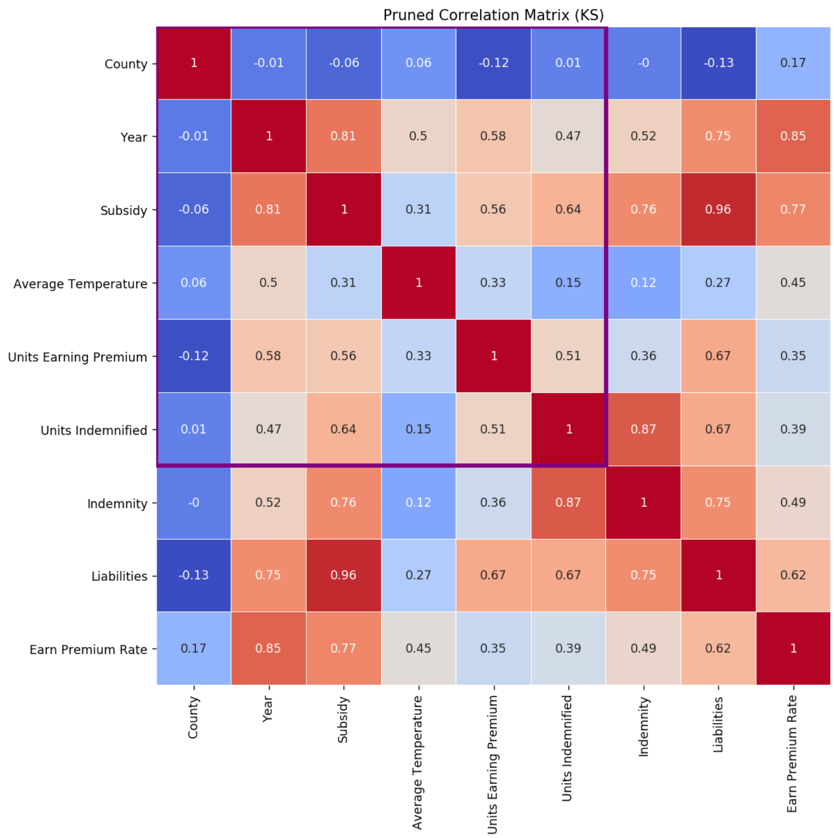
- Subsidy vs. liabilities (outside the box)
- Earn premium rate vs. total premium (outside)
- Indemnity vs. units indemnified (outside)
- Indemnity vs. subsidy (outside)
- Earn premium rate vs. subsidy (outside)
- Policies indemnified vs. loss ratio (outside)
- Subsidy vs. year (inside the box)
- Total premium vs. year (inside)
- Earn premium rate vs. year (inside)
- Policies sold vs. policies earning premium (inside)

We see that much of the data in the detailed heatmap has multicollinearity, which occurs when two features have a warmer color within the purple box; this means that they are too closely related. In a broader context, multicollinearity is the existence of near-linear relationships among the independent variables, where one variable can be used as a good predictor or indicator of another. By reducing multicollinearity, we in turn reduce the number of variables while still maintaining roughly the same amount of precision, accuracy, and information in the model.

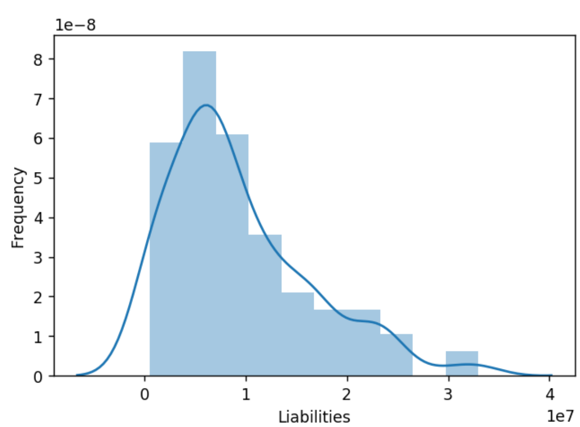
In order to check for multicollinearity, we must check for closely related features inside the purple outline, as mentioned before. We should not select both of these features together for training the model; one must be eliminated. To understand the reason multicollinearity is harmful for regression, consider the following scenario with features A, B, and C. If A is regressed against B and C together, where B and C are highly correlated features, then it is extremely difficult to distinguish the individual effects of B on A from C on A, because any increase in B tends to be associated with an increase in C. In summary, multicollinearity makes it hard to assess effects of certain independent variables on dependent variables. Therefore, to counter this problem, we must create a simplified heatmap with a smaller number of features that still explain the data well.

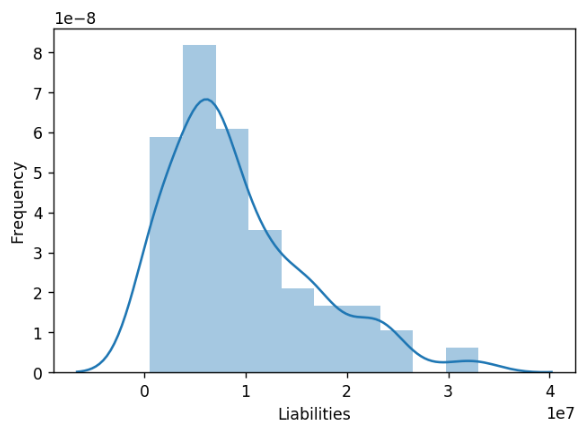
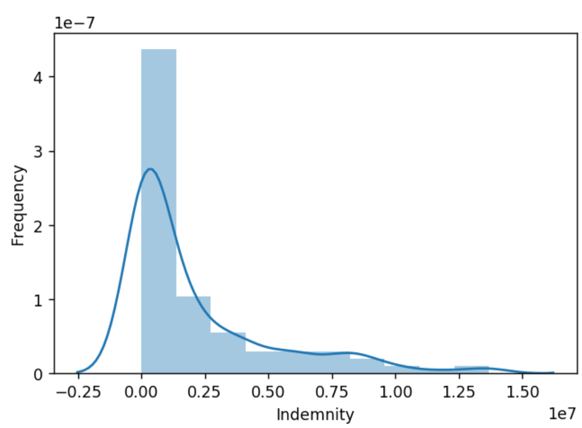
Given on the next page is the result of removing features that suffered from multicollinearity. We note there are high magnitude values in cells where year is involved, which means time itself captures a significant portion of the variation in the data. We leave the county as a feature in this overarching regional model despite its associated low correlation coefficients with other features, in county turns out to be a more significant player for other regions besides southwest Kansas (e.g., northeast Washington).

Essentially, this heatmap encompasses single possible regression of insurance-related features and targets as identified previously. Because we have addressed multicollinearity, we see mostly light (low magnitude, either positive or negative) in the grid squares contained within the purple outline. For example, either one of “Policies Earning Premium” and “Policies Sold” was removed, because they have a 0.98 relationship. The features **County**, **Year**, **Subsidy**, **Average Temperature**, **Units Earning Premium**, and **Units Indemnified** are the six features that were used in the final regression to predict target factors, **Liabilities**, **Indemnity**, and **Earn Premium Rate**.



Next, three insurance-related models were created: one to predict liabilities, one to predict indemnity, and one to predict earn premium rate. The three models were trained on the six listed features and learned weights for each feature to determine the best possible expression to predict indemnity, liabilities, and earn premium rate. The data from the features were split randomly into 80% for training the model for prediction and 20% for testing the model, and their respective R^2 values are shown below. We can manipulate these models using the formulas given previously to get the desired target, which is loss ratio. In the table below, we provide distributions of the targets, as well as the coefficients of the multiple linear regression models, which when written in equation form, can be used to predict target values into the future, e.g. until 2050.

Target Factor	**Graph of Model**	**Results**
Indemnity		<u>Coefficients</u> County: **16609.607141** Year: **-14215.087883** Subsidy: **1.565381** Average Temperature: **-103141.154103** Units Earning Premium:

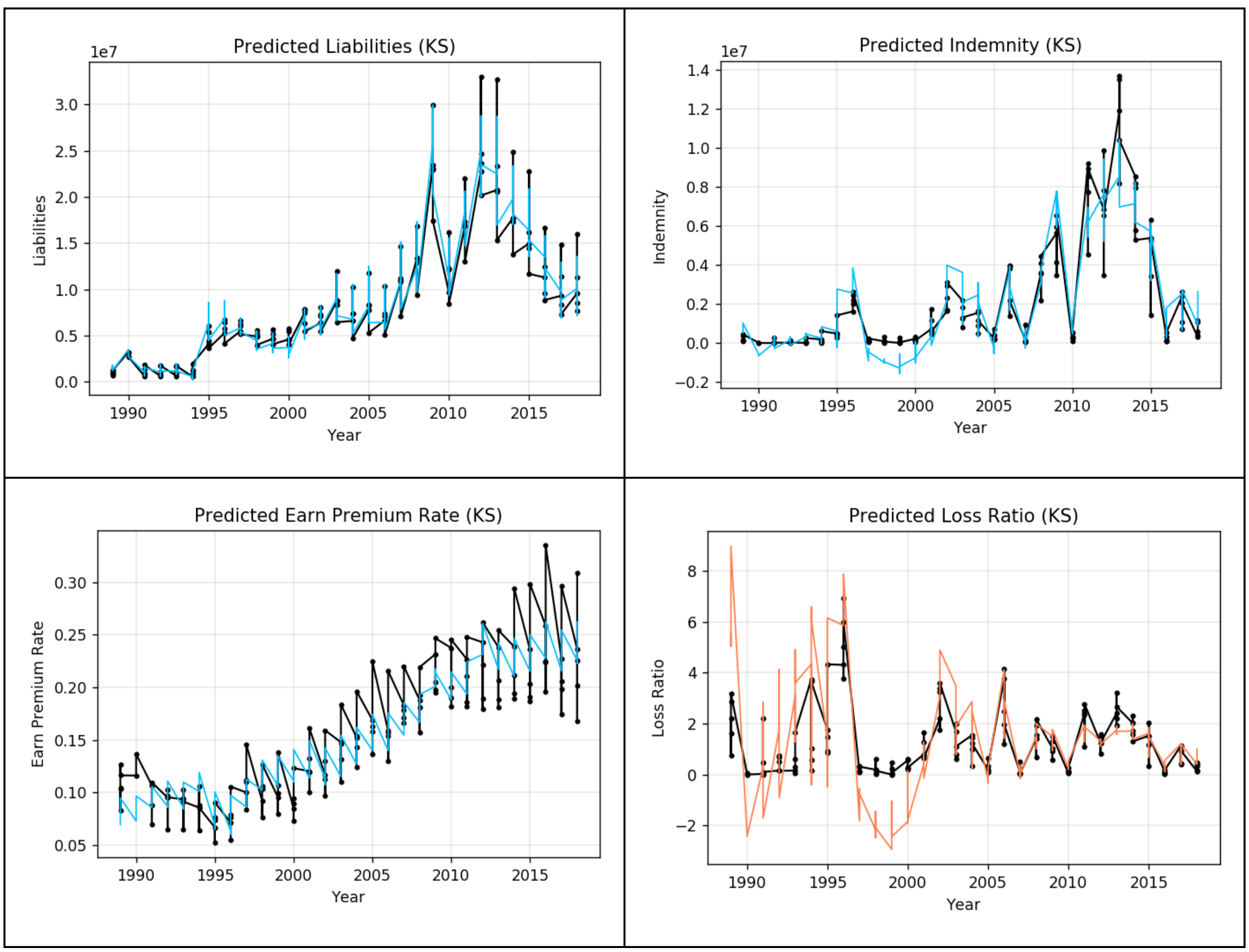
	R^2 Values, RMSE: <u>Training</u> 0.869237056906817, 1062690.8259947817 <u>Testing</u> 0.8859604048287234, 1097662.2932482888	**-1338.114398** Units Indemnified: **4686.457999** Intercept: **34169922.65774365**
Liabilities	 R^2 Values, RMSE: <u>Training</u> 0.9606728325395607, 1407743.491926344 <u>Testing</u> 0.9708209536500987, 1195184.2640275895	<u>Coefficients</u> County: **-192980.005776** Year: **-110655.082286** Subsidy: **7.166490** Average Temperature: **-154915.051076** Units Earning Premium: **2829.062541** Units Indemnified: **385.544901** Intercept: **2.29464558e+08**
Earn Premium Rate	 R^2 Values, RMSE: **Training** 0.8283887570134867, 0.025998137397301974 **Testing** 0.7893974590450333, 0.0268934197146078	<u>Coefficients</u> County: **6.878908e-03** Year: **4.772891e-03** Subsidy: **2.734039e-08** Average Temperature: **3.709994e-03** Units Earning Premium: **-2.487706e-05** Units Indemnified: **-1.291524e-05** Intercept: **-9.62492127**

The closer the R^2 value is to 1, the better the trendline fits the data and explains variations in it. We always want an R^2 much closer to 1 than 0. Also, it is shown that the model usually performs better on the test set of data, which consists of values that the model tries to predict without having seen them during training. Given the comparable performances of the three models, it is possible to conclude that the models are not extremely biased and are able to generalize and find a global trend in the data in order to predict future values.

We can also evaluate the performance of the models using RMSE, or the root mean squared error, which is the standard deviation of the prediction errors of a model. There is not a certain fixed threshold for good RMSE and R^2 values, but it is usually best to have the RMSE value as low as possible. If RMSE of the training data is less than the RMSE of the test data, then the model is overfitted; overfitting refers to the idea where an analysis corresponds too exactly to a particular dataset and therefore fails to predict future observations with high accuracy or reliability. Underfitting, which is the opposite of overfitting, can occur if the RMSE of the training data is greater than RMSE of the testing data.

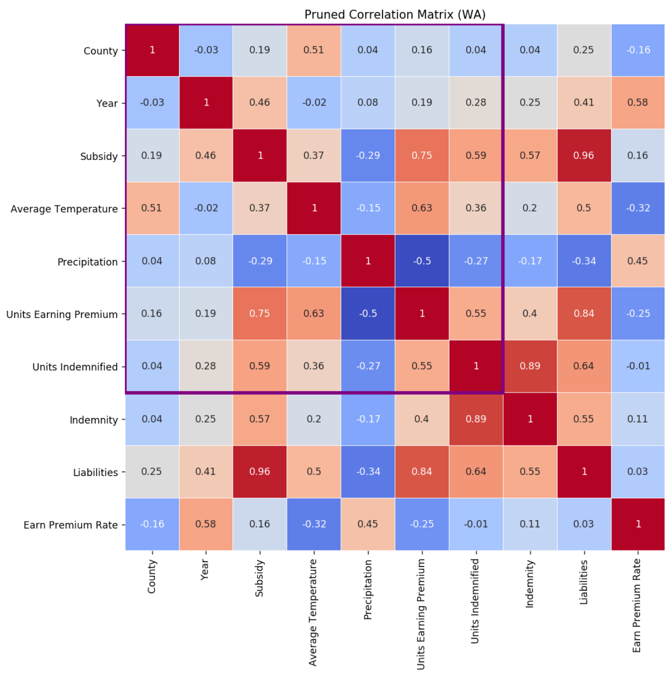
The third column of the table above displays all of the 6 coefficients found by the multi-regression model, as well as the intercept of the line of best fit. In equation format, this would appear as:

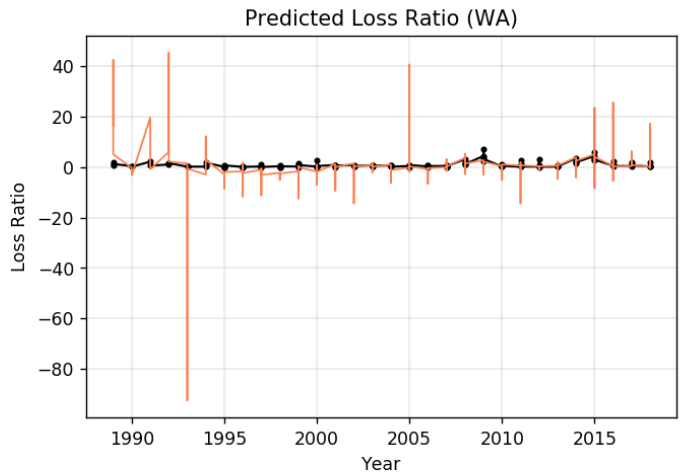
$$y = 16609.607141A - 14215.087883B + 1.565381C - 103141.154103D - 1338.114398E + 4686.457999F + 34169922.65774365$$



For the graphs above, all five counties for Kansas were consolidated into one plot per factor (liabilities, indemnity, earn premium rate, which are combined to predict loss ratio), and each county is represented by points that are stacked vertically for each year. The colored lines in the plots show the predicted trendline, which can be continued into the future as given by the generated regression equations. For generating predictions into the future, e.g. up to 2050, we ran individual regressions for features over time, and then plugged those into the regression, with year and county as known inputs. Several features correlated well with time, so these sub-regressions, and thus the model itself, performed well.

Then, similar procedures for Washington were performed. One additional feature survived the pruning process, namely precipitation. Below are the results of the predicted loss ratio graph.



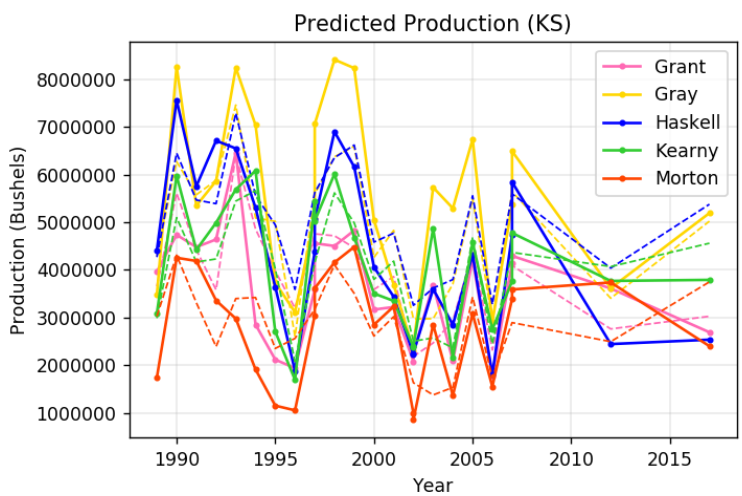


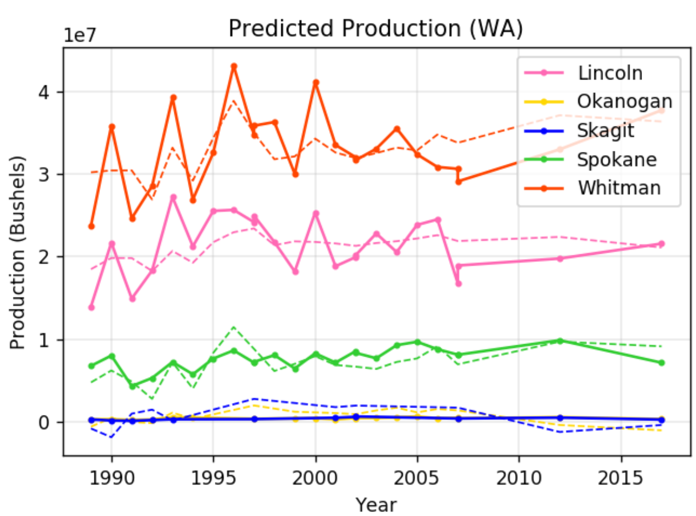
It seems that, in predicting loss ratio, the Kansas regional model performs better than the Washington regional model despite the same methodology applied primarily due to the fact that they were so different in their makeup, as it captures trends more closely without overfitting, as evidenced by the RMSE values of the loss ratio's components (for which all RMSE train and test values are about the same). Evidently, geography and agricultural products are not the only factors to consider when applying, creating, and testing a model. There are other factors that account for the performance differences — we examine the effect of climate on wheat yield to see if differing temperature, precipitation, and drought patterns contribute.

Taking wheat production (in bushels), average temperature, precipitation, and drought indices as features, we trained a second-degree polynomial regression model to achieve the following, where it was deduced that a polynomial regression was better than a linear one since the RMSE values decreased while the R^2 score increased. The Washington region's agricultural yield forecast model performed better this time, meaning that it is likely that Washington's subpar performance in the loss ratio prediction task was more due to other relative feature differences between counties of the region than due to climate-associated differences.

KS	RMSE, training: 961871.3840827306 R^2 score, training: 0.63370456832612 RMSE, test: 1043056.8367301229 R^2 score, test: 0.7489272039361641	coef:[0.000, -11948710.844, -33045463.199, -5999729.247, 39314739.467, -99618142.191, -419451.712, 6290.199, 31120.933, -6396.241, -12726.403, 8194.142, 9872.218, -20659.026, 52773.354, -128418.357, 48731.662, -126137.951, -18075.008, 88556.145, -44009.685], intercept:32942003129.26
WA	RMSE, training: 2423422.8953938624 R^2 score, training: 0.9662501668849458 RMSE, test: 2501298.8895306317 R^2 score, test: 0.9583467834731139	coef:[0.000, -54926601.638, 35408356.915, -45348407.502, 554809.058, 81395776.852, 8660228.356, 8843.197, -276904.745, 92744.303, -63729.601, -9075.567, 22099.905, -912.293, -42325.494, 9841.458, 29507.097, 72415.196, -2568.402, 2398.513, 101837.967], intercept:-34421356747.51

The production plot trends are given in dotted lines, while the ground truth data is solid:





Risk Analysis

Based on the predicted production graphs of Kansas, wheat production has not been stable within the last three decades, and there is no guarantee that it will become stable in the near future. Many of these fluctuations are due to climate change, because adverse weather often makes yields more variable from year to year, causing losses to occur more frequently. Earn premium rates as well as premiums, are steadily on the rise (as implied by a positive slope), which pose a risk to farmers in this industry. The increased probability of a climate-related disaster accompanied by escalated prices to be paid for insuring crops threaten the financial success of farmers.

Indemnity, as shown in the indemnity prediction plot in the Results section, clearly depicts a strong positive correlation, indicated by the high R^2 value of about 0.886 (the probability that a high indemnity will occur in a given year). Similarly, liability is shown to increase as well with an R^2 value of 0.971 (the probability that a high liability will occur in a given year). This means that wheat agricultural production will likely see losses in the future, and insurance companies, policymakers in the government, and farmers must take precautions to address this. At current rates, wheat production, because of climate change, is declining at a rate of 0.9% per year, which equates to millions of bushels of wheat lost in the US annually.

There are also several industries that will be affected by changes to wheat production in the agricultural industry. Risks, primarily involved in the energy, livestock and poultry, and snack food industries, will greatly affect people, the community, and the economy of the government.

<u>Energy:</u> Due to advances in technology and its great sustainability, using wheat products is now as efficient and beneficial as using other commodities such as corn or soybeans. Wheat has a wide range of uses, because it is so easily produced, and the necessary production and distribution methods have already been settled.

Because of technological advances, countries have begun to use wheat in the production of ethanol, gas, and oil products. While the US primarily utilizes corn for this purpose, wheat is a much more sustainable resource for the environment. According to the International Grains Council, Canada, China and the European Union (consisting of a large group of countries), use wheat as the principal grain for making ethanol, since corn is not as widely available as it is in the US. Of the 7.6 million tons that the E.U. will process into ethanol this year, wheat accounts for 60 percent, followed by maize and then barley.[13]

However, use of fossil fuels will continue to become more restricted. When combined with potential threats of reduced wheat production, the lucrative oil, gas and coal industries are at risk. “The movement away from fossil fuels will have a big impact which could affect banks and investment firms that have relationships with the fossil fuel industry,” said Geoffrey Heal, a professor at Columbia University’s Business School. He comments that the stock market value of the US coal industry in 2011 was about $37 billion, but it massively declined through 2019, and is now worth about $2 billion.[14] People and companies that made large investments in these industries could suffer in debt, and they have suffered major job cuts.

<u>Livestock and Poultry:</u> Wheat is the premier grain for feeding livestock and poultry, and its successful harvest is now challenged by climate change and industrial uses.[12] The latter is much

more minor, but as wheat begins to play a larger role in forming the basis of ethanol and oil products, the amount of wheat available for feeding lowers slightly. However, the much more impactful factor is climate change's effect on wheat.

As with many other cereal grains, wheat is packed with carbohydrates and is a stable source of energy. It is usually preferred as an energy source to other grains such as corn because of its higher available energy, expressed as the amount of digestible or metabolizable energy per unit of dry matter.[12]

Low-quality meat that is unsuitable for milling (because of pests, weather damage, etc.) are usually fed to animals. It is worth less than good quality wheat, but it still contains all necessary vitamins and benefits for animals.[12] Thus, wheat is thus a major component of animal feed. The reduced harvesting of wheat poses a threat to the poultry industry. Wheat is such a large component of many animal diets; without it, or with a reduced amount, the amount of available animals raised successfully would decline. This in turn affects many economies who rely on meat and the people who consume it.

<u>Snack Foods:</u> Wheat is an ingredient in an overwhelming portion of snack foods like cookies, crackers, and pastries. Unfortunately, the grain-based foods industry is losing market share to other food categories, and this can be partially credited to the consistently declining production of wheat. Volumes of many major wheat-based foods have been seen declines in recent years; most concerning is the flagship product — bread. Bread volume sales reduced by 4.8%, while dollar sales experienced a 0.8% decline as well.[15]

Food Business News provided a list of other wheat-containing foods that had reduced sales in 2019 compared to 2018.[15]

- Hamburger and hot dog buns: -2.4%
- Other rolls: -1.8%
- Pita bread: -7%
- Crackers, crackers with filling, saltine crackers: -4.2%
- Graham crackers: -3.1%
- Breadsticks: -8.3%
- Ready-to-eat cereal: -3.1%
- Popcorn: -0.8%
- Hot cereal: -4%
- Pretzels: -3.9%
- Granola bars: -5%
- Pasta: -0.7%

The potentially worsening situation of wheat production will only have further negative effects on this important industry in the years to come.

Recommendations

The vulnerability of agriculture to climate change is strongly dependent on the responses taken by humans to moderate the effects of climate change. Adaptive actions within agricultural sectors are driven by perceptions of risk, direct productivity effects of climate change, and by complex changes in domestic and international markets, policies, and other institutions as they respond to those effects within the United States and worldwide. Opportunities for adaptation are shaped by the operating context within which decision-making occurs, access to effective adaptation options, and the capacity of individuals and institutions to take adaptive action as climate conditions change.

Effective adaptive action across the multiple dimensions of the U.S. agricultural system offers potential to capitalize on emerging opportunities and minimize the costs associated with climate change. A climate-ready U.S. agriculture will depend on the development of geographically specific, agriculturally relevant, climate projections for the near and medium term; effective adaptation planning and assessment strategies; and soil, crop and livestock management practices that enhance agricultural production system resilience to climatic variability and extremes.[16] New research and development in new crop varieties that are more resistant to drought, disease, and heat stress will increase the resilience of agronomic systems to climate change and will enable exploitation of opportunities that may arise.[16]

Droughts are the most prevalent cause of loss in the dataset. Hopefully, with more focus on how to best combat this climate change issue, governments will be able to initiate action and policies to aid suffering locations. Initiating weather derivatives, which are instruments used by insurances and financial institutions to hedge against the risk of the risk of weather-related losses, will be a good strategy for managing the effects of droughts.[17] Typically, they are used to help farmers combat the risks of poor harvests due to unexpected hot or cool temperatures or other climate changes. Insurance companies can employ weather derivatives to smooth farmers' earnings and maintain their consistency when climate and temperature hinder production.

Since liabilities are likely to become more frequent, severe, and volatile, insurance companies should change, for example, some of their long-term bonds in high-risk areas to short-term bonds in an effort to conserve money should a loss occur. A second strategy is to invest funds in assets that increase in value during drought periods (because drought is the most prevalent cause of loss in Kansas). For example, this could be investing in stock from irrigation system manufacturers. This would help lessen the blow of a dramatic loss.

At the sector level, governments should ensure the availability and distribution of relevant information on using and conserving resources efficiently, and they should initiate more public policies and risk management to aid farmers and make informed investments in adapting to climate change. Strengthening access to knowledge and risk management methods is vital to increasing adoption of sustainable and productive practices.

Lastly, as advice to farmers, where possible, harvesting wheat and other crops should be geared towards sustainability. For example, farmers could plant different crops within a single region to create more biodiversity in their fields, rather than planting single large sections of the same crop. Currently, it is more important to focus on the highest-yielding methods rather than the most efficient ones; however there still must be a balance of both.

References and Citations

[1] Shewry, P. R., & Hey, S. J. (2015). The contribution of wheat to human diet and health. Food and Energy Security, 4(3), 178-202. (link)

[2] Kansas State Research Shows Climate Change Impact on Wheat Production. (n.d.). (link)

[3] Gander, K. (2019, September 25). Climate change predicted to hit wheat production-and the U.S. is going to be one of the worst places affected. (link)

[4] Kiprop, J. (2018, November 29). US States That Produce the Most Wheat. (link)

[5] Climate Change and Agriculture in the United States: Effects and Adaptation. (2013, February 6). [PDF file]. (link)

[6] 2100 Warming Projections. (2019, December 10). (link)

[7] Kansas AgGrowth: 2019 Wheat. (2019). [PDF file]. (link)

[8] All About Grains: Wheat. (2019, July 15). (link)

[9] Campbell, M. (2018, November 26). Midwest farms may suffer huge losses from intense heat, says climate change report. (link)

[10] What Climate Change Means for Kansas. (2016, August). [PDF file]. (link)

[11] Preparing for a Changing Climate: Washington State's Integrated Climate Response Strategy: Agriculture. (2012, April 3). [PDF file]. (link)

[12] Wheat for Animal Feed. (2020, January 24). (link)

[13] Lyddon, C. (2017, March 20). Industrial uses for grain expanding. (link)

[14] Industrial use of grains continues rising. (2014, June 3). (link)

[15] Sasland, M. (2010, May 4). Don't rule out wheat as industrial use expands. (link)

[16] Gardner, L. A., & White, T. A. (2014). Managing Investment, Underwriting, and Production Risks from Drought-Related Agricultural Exposures. [PDF file]. (link)

[17] What are weather derivatives? (2016, December 1). (link)