Module 2: CONTROLS AND AUDITING

Table of Contents

Module 2 Information

Controls and Auditing: Introduction

[MUSIC]

This module introduces you to important concepts related to internal and external auditing. It also introduces you to the concept of internal controls.

First, the module will provide an overview of the technology related to the audit function within accounting. Next, the module will discuss the financial reporting environment and current pressures related to that reporting environment, including both earnings management and outright fraud in financial reporting. Further, the module discusses the fraud triangle and fraud and earnings management from upper level management such as CEOs and CFOs and also from lower level employees. The module will then continue by discussing the concept of internal controls and the need to test those internal controls and how those interact with the financial reporting environment. The module concludes by introducing and solving a realistic case about auditing the internal controls around procurement cards. Also known as P cards. P cards are like credit cards that businesses provide to their employees for use and business purchases. The case uses real data from P card transactions from Oklahoma State University. First the case brings in the data and cleans it. Next it explores the data and finally test three specific controls related to the use of P cards at Oklahoma State University. The case will be done in both Alteryx and Rstudio. As always, I encourage you to follow along closely and work through the case yourself in order to make these skills a valid part of your data analytics toolkit.

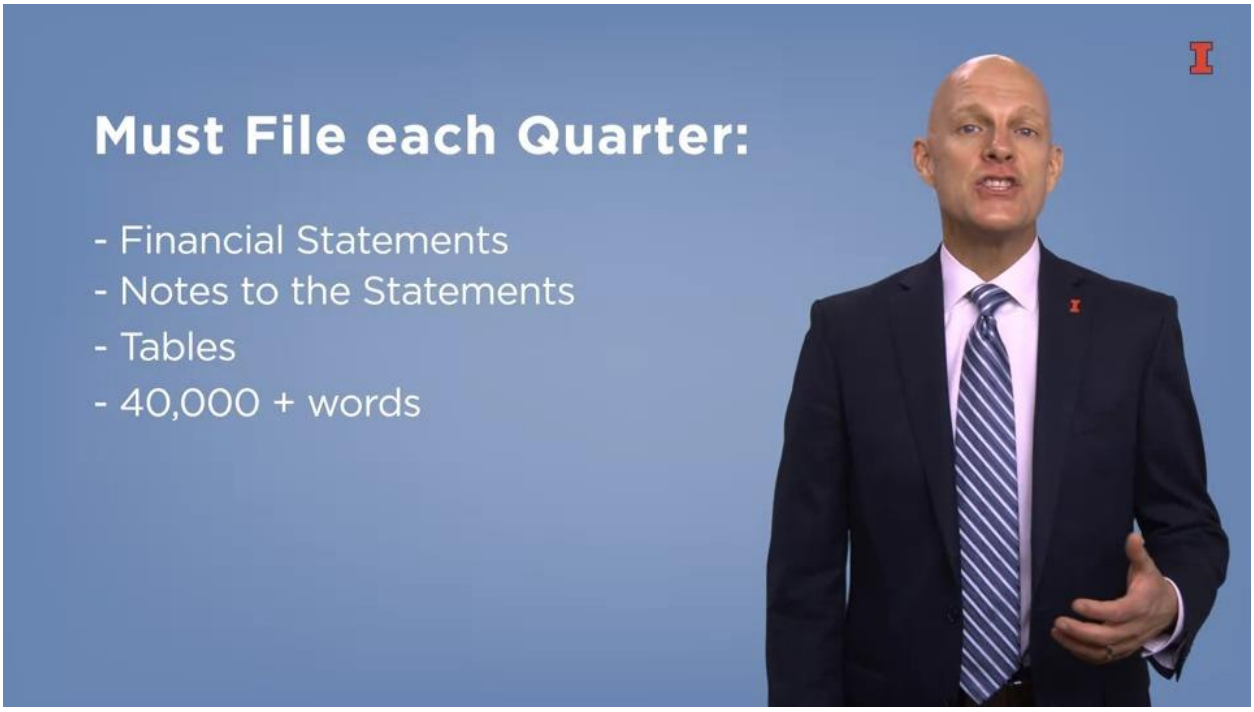
Module 2 Lectures

Lesson 2-1: Audit Technology Overview

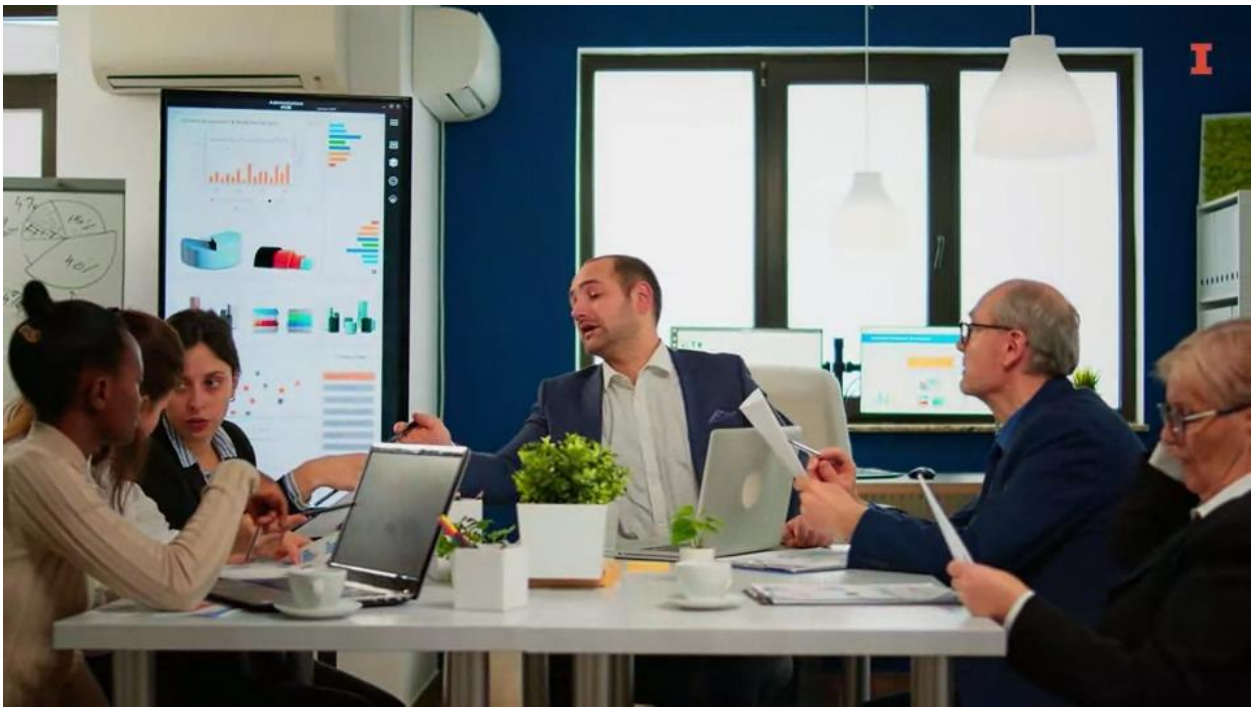
[MUSIC]

In this lesson, I'll provide an overview of advances and technology in financial auditing.

Recall that public companies that are owned in whole or in part by investors in the general public, must file their financial results with the government.



Government regulation in the United States requires that these public companies file financial statements, other notes and tables related to the financial results and risks of the company each quarter. These documents are extensive, for example, the larger yearly submission called the 10K, is usually over 40,000 words and can be as large as 100,000 words.

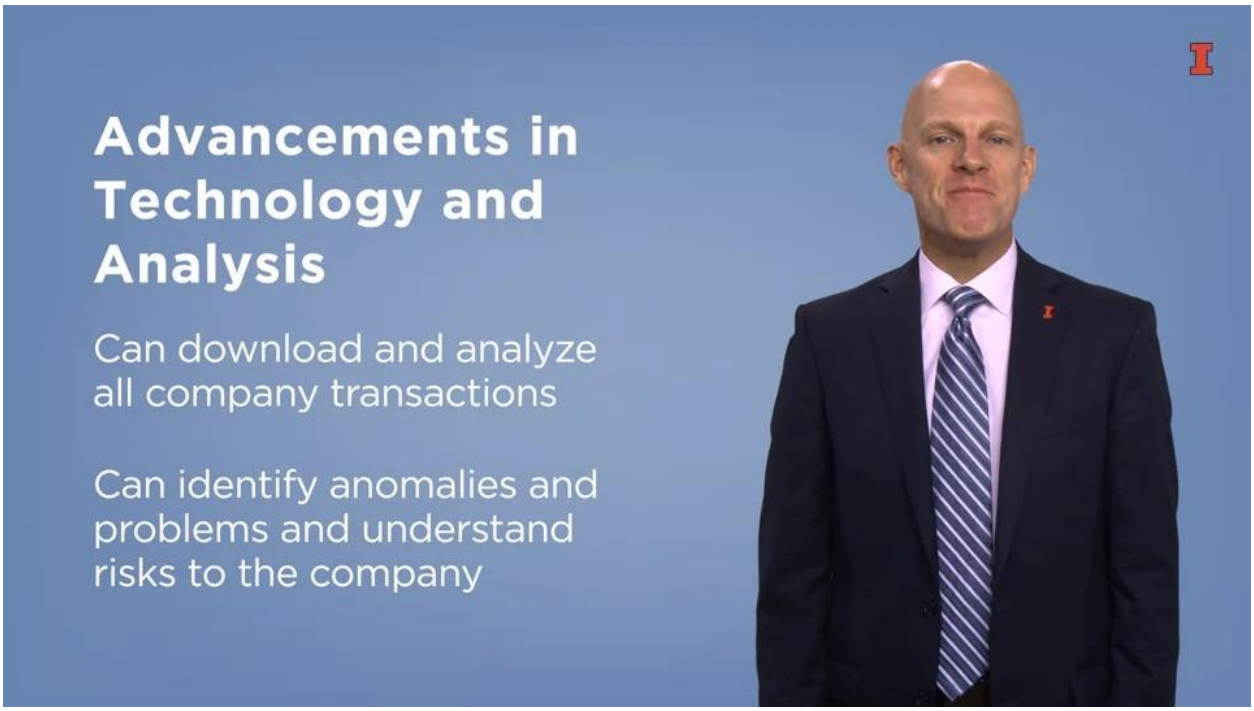


Financial auditors fill the government mandate to review public company financial statements, and ensure that the accounting in those reports follows the government rules, at least to a material respect.

Auditors speak to the management of a company and examine enough of the financial statements so that they can issue an opinion that the company's financial accounting follows the rules. These tasks include such things as reevaluating and recalculating management's estimates. Inspecting the company's inventory using data analytics and visualization to test all of the company's transactions. Contacting banks and customers of the company to confirm that they agree with the balances that the company reports on the financial statements. Inspecting physical assets that the company lists as assets. And inspecting individual inputs that account for the revenue that the company claims to make sure it's listed following the rules.

The auditor uses several different types of technology to increase the effectiveness and efficiency of the audit.

These include the use of mobile and cloud technologies, robot process automation, Blockchain process mining, artificial intelligence, machine learning and even drone technology.



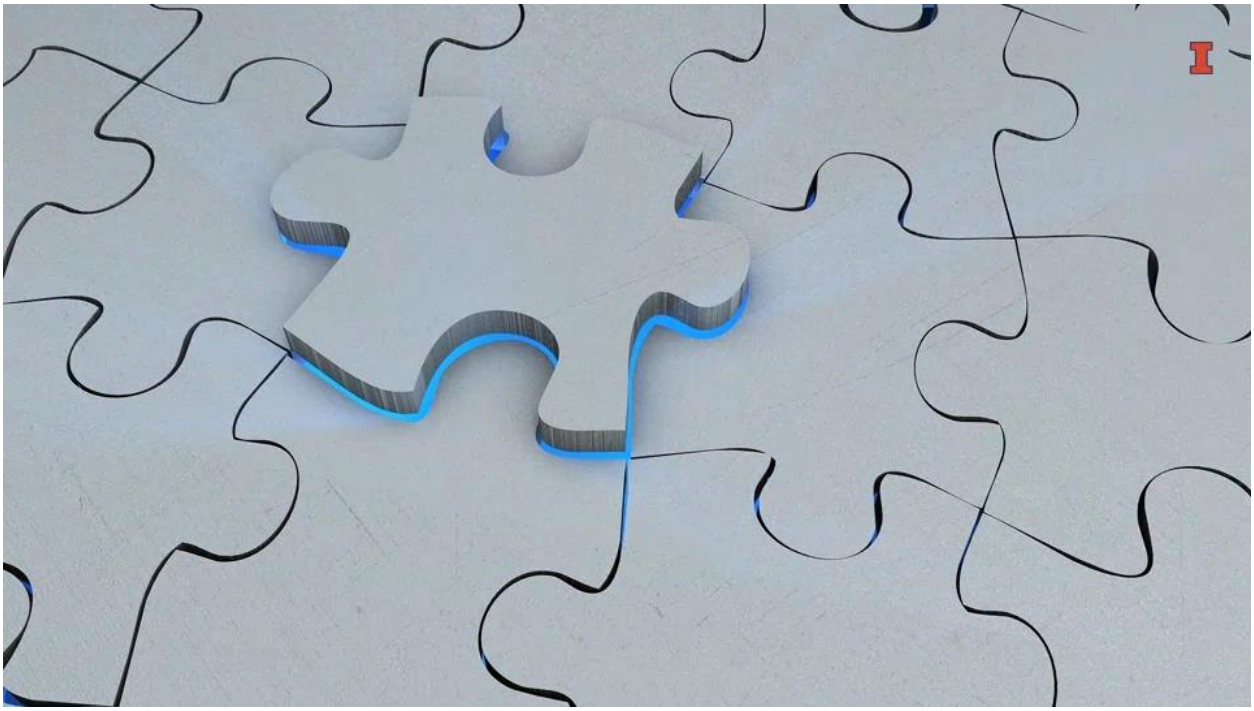
Before recent advances in technology, the only way auditors could test management's results and numbers was to sample the results. That is, archers would pick a few of the thousands or even millions of transactions comprising the financial statements of the company and analyze just those few, and then use statistics to make guesses about the full population of the company's results. However, this testing approach, while effective, is somewhat flawed, and that it can only give a statistically proximate tests of the results. For example, consider a company whose primary business is selling software as a service, or selling convenience store items. These companies have many, many transactions that are low in value. Thus, when the auditor audits revenue, they'll have millions of revenue transactions to test that are all under $20 apiece, for example. Thus with sampling, the auditor would have to select a huge amount of revenue transactions in order to get any amount of comfort that the total revenue amount that is the revenue balance is correctly stated. Now, with advances in technology, the auditor can download 100% of the company's transactions and conduct analysis on each one of those revenue transactions. Now, instead of having to only examine a few of the transactions, the auditor can download the entire population of transactions, and trace each sale forward from the point of the sale to the collection of cash from the customer, and backward from the point of sale to the purchase of inventory. Thus, auditors can now do an even better job of letting the data speak. Analyzing the client's full set of transactions instead of just a part, allows the audit team to identify trends to better understand the source of the transactions and how transactions flow within the company, and to better understand how different accounts correlate and move together. This allows the auditor to identify anomalies and problems, and to better understand the risks the company faces and the risk that the financial statements might be misstated.

Of course, getting the data from the clients and getting it into a format that makes it easy to analyze is no easy feat. Most data analysts will tell you that approximately 80% of their jobs are extracting, transforming and loading data or ETL, before the data is even able to be analyzed. This is no different for the auditor. Thus, the first steps after extracting the data are transforming it, cleaning it and loading it into the necessary tools so that it can be analyzed. Thus, like any data analyst, the auditor should understand the format of the data, what it looks like and what it should look like. The auditor also needs to understand common errors with data extraction and transformation, such as merge or join errors, missing data, etc. While some of the largest audit firms have shared services that will do these tasks for the auditor, it still behooves the auditor to understand the data, and what it should look like, and to be able to perform these practices for themselves.

At the very least, auditors who have this understanding are more likely to be successful in their.

Capturing all of the data for an audit client is the first step. The next step is analyzing that data. This analysis can prove vital to helping the auditor identify potential areas where the financial statements might be incorrect. For example, these analysis could help the order to identify employees who are charging more hours than is reasonable. In another example, these analyses could detect and analyze historical trends to predict future profitability of products. I'll discuss two parts of this.

First, the auditor benefits greatly from visualizing the data.



A picture is truly worth 1000 words as the old saying goes,

and visualize analysis of the data allows the auditor to spot trends and potential problems much more quickly. Second, audit analysis often creates a score for each transaction that indicates how risky or problematic that transaction is. For example, if the transaction is entered manually by someone who normally does not enter transactions, this transaction would be flagged as problematic. Other examples include having new transactions that happened when most stores are closed, such as holidays. The auditor can use past history with that company and machine learning to generate this risk score.

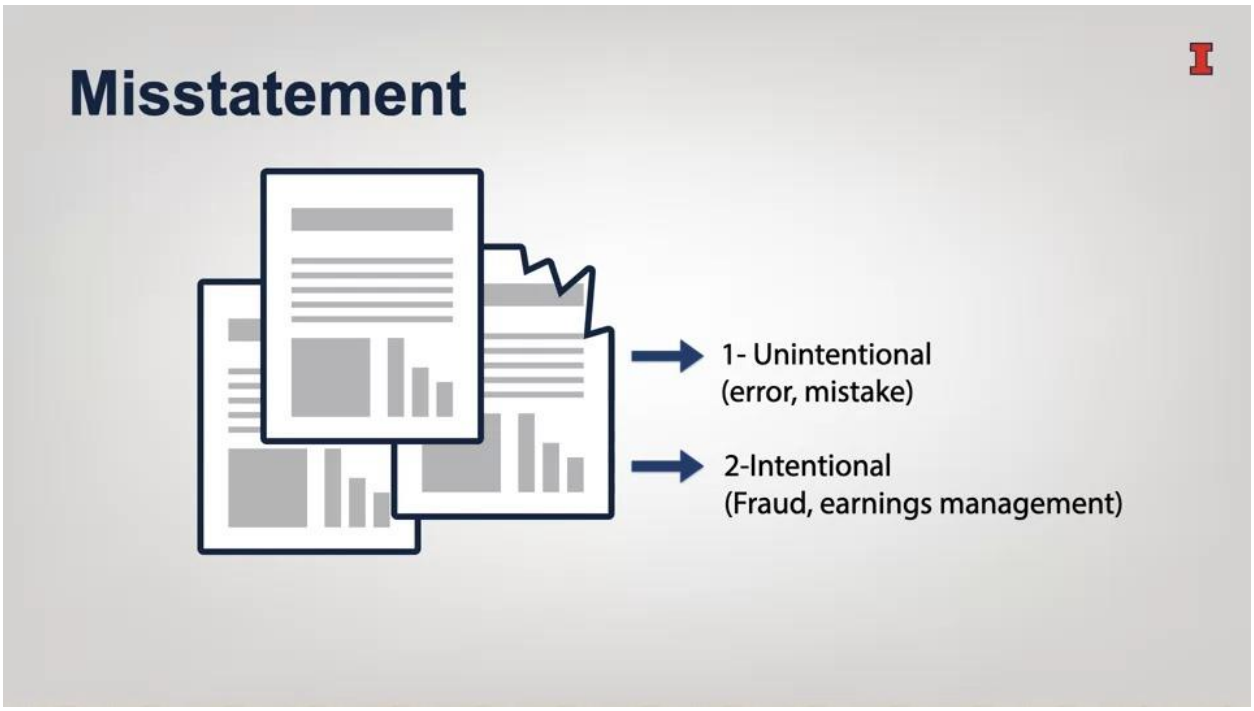
Two recent auditing leaders were quoted as saying,

as we look ahead, data and technology are clearly the keys to high quality auditing, which promotes high quality financial reporting and gives investors trust and confidence in the capital markets.

Lesson 2-2: Reporting Pressures, Earnings Management and Reporting Fraud

[Music]

The job of financial auditor is to make sure that the financial statements presented and reported by a company their audit client follow the rules for generally accepted accounting principles.

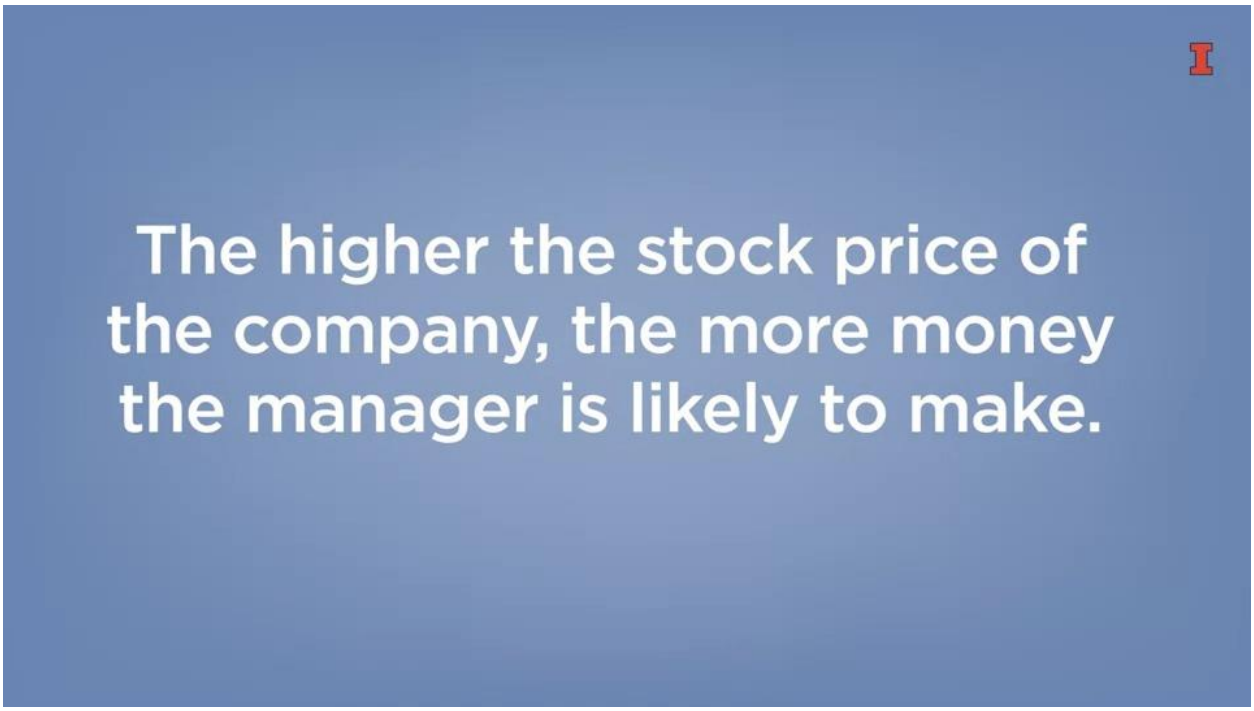


This means that the auditor is looking for mistakes which are usually called misstatements. These mistakes can be both intentional and unintentional. This video focuses on intentional misreporting.

Businesses, especially public companies that are owned in whole or in part by investors in the general public are under significant pressure to perform well.

Public companies are the large companies that sell their stock to the public and thus must file their financial results with the government.

This pressure can sometimes lead companies to lie, cheat, and commit fraud. Managers who are paid to run a company want the company to be highly valued by stockholders and potential stockholders and bankers. There are two main reasons for this. First, when a company that you run is doing well it reflects positively on you. Thus, you're likely to retain your job, receive more compensation and be sought after for even more prominent and lucrative jobs. Second, managers very often receive financial incentives when the companies they run are successful. For example, many managerial bonuses are tied to performance and many managerial financial incentives are tied to the price of the company stock.



Thus, the higher the stock price of the company, the more money the manager is likely to make.

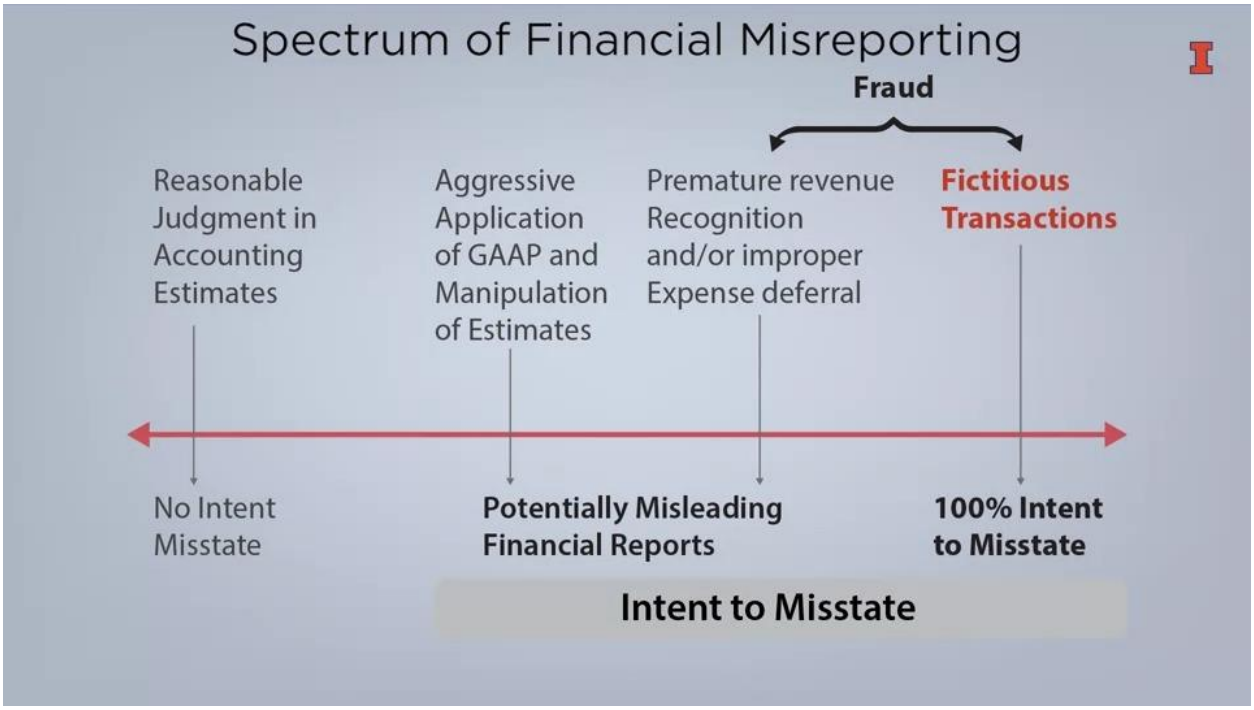
These incentives while helpful and encouraging the managers to maximize the value of the company, can also add perverse and negative consequences. Specifically, managers whose companies are not performing well have an incentive to overstate and even to lie about their company's performance to make it seem better than it actually is. Thus, the reported accounting numbers may be manipulated, managed, and fraudulent. This is generally called earnings management or fraud.

These pressures that managers feel are particularly acute when financial profitability is threatened by economic industry or entity operating conditions such as a high degree of competition, declining margins, high vulnerability to rapid changes such as changes in technology, product obsolescence or interest rates, a declines in customer demand, operating losses, making the threat of bankruptcy foreclosure or hostile takeover imminent, recurring negative cash flows from operations,

rapid growth, especially compared to the same industry and new accounting, statutory or regulatory requirements. Managers might also feel pressure to meet the requirements of third parties. Some examples of these pressures are profitability expectations of analysts and institutional investors. They need to obtain additional debts or equity financing, meeting exchange listing requirements or debt repayments or other debt covenant requirements. Perceived or real adverse effects of reporting poor financial results on significant pending transactions, such as a business combination or a contract awards. Finally, management might feel pressure due to their own personal financial situation, such as significant financial interests in the entity.

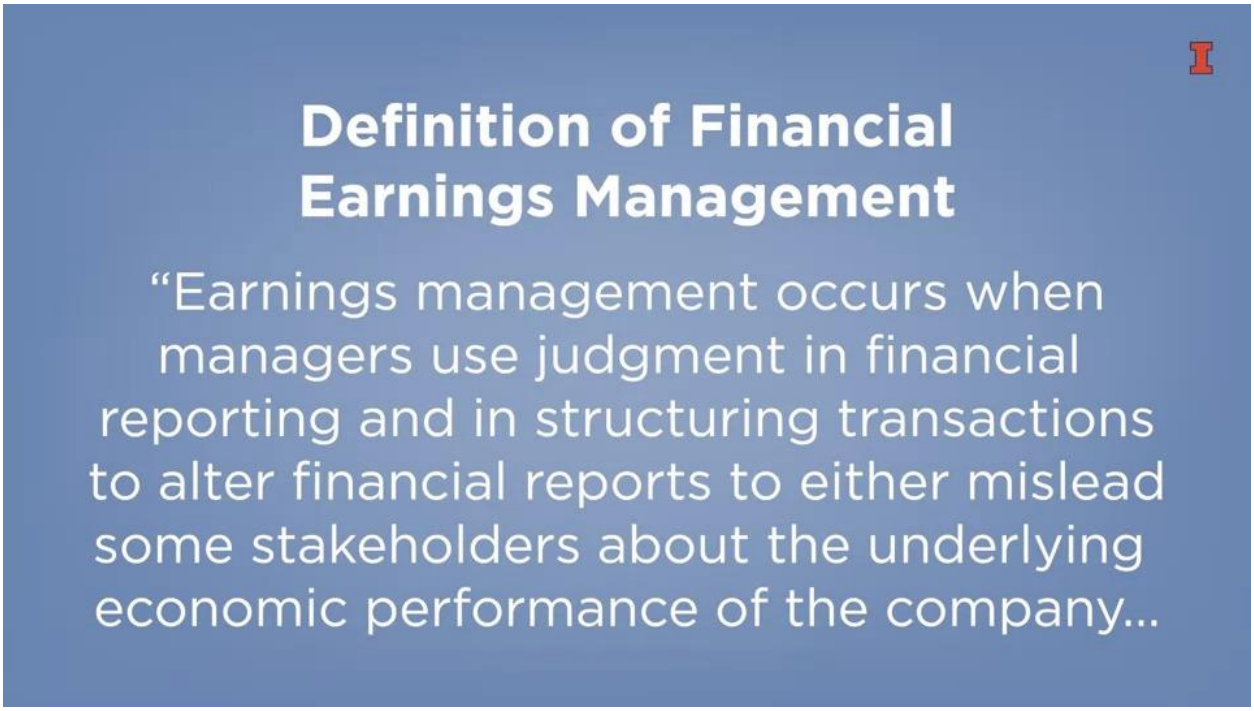
Significant portions of their compensation like bonuses or stock options, are earned on arrangements and being contingent upon achieving aggressive targets for stock price, operating results, financial position or cash flow. Personal guarantees of debts of the entity. As one example the Chair of Graham Harvey in Raj Cupola find that a large majority of CFOs feel that earnings misrepresentation occurs most often in an attempt to influence stock price because of outside and inside pressure to hit earnings benchmarks and to avoid adverse compensation and career consequences for senior executives.

Managers who succumb to these pressures and cheat to make their financial results look better than they really are, have many avenues to do so.

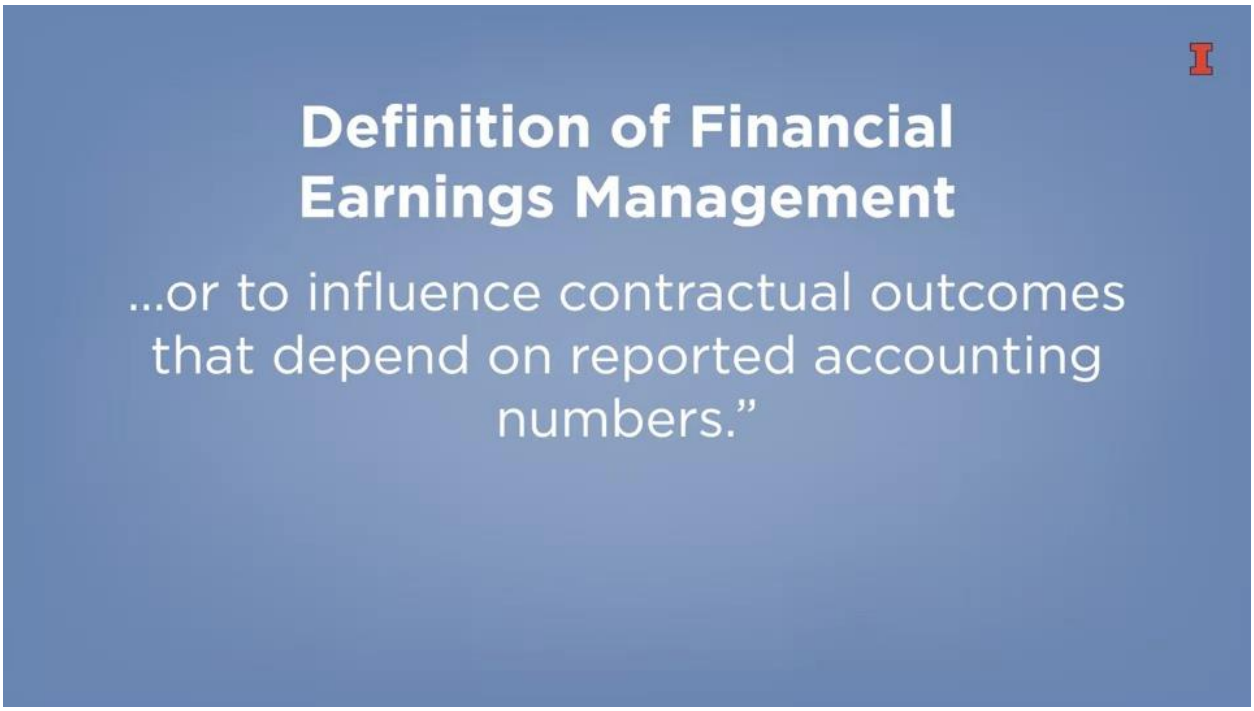


Take a look at this spectrum of financial reporting. On the far left, we have a situation where the financial reports are 100 percent accurate to the best judge of management. On the far right, we have fully fictitious transactions. At this extreme, managers might commit fraud whereby they generate 100 percent fictitious transactions, such as recognizing revenue that doesn't even exist. These misstatements are clearly illegal. Now, if we move back a little to the left on the spectrum or towards the middle, we see actions that managers might take such as recognizing revenue before they have actually earned it, or improper and delayed expense deferral. These actions may or may not be strictly illegal. Moving further to the left, managers might engage in something called earnings management. Here, managers don't explicitly break laws or even accounting rules necessarily, but take advantage of the rules of generally accepted accounting to report as aggressively as possible.

A definition of financial earnings management is as follows.

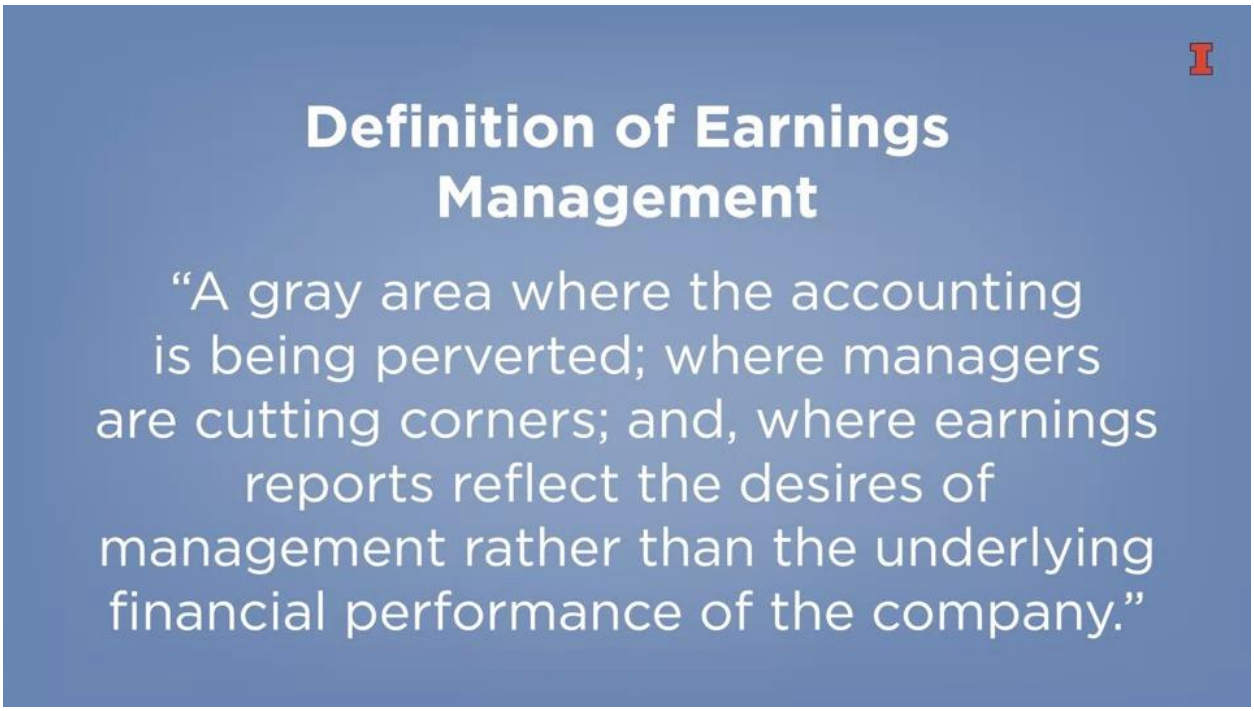


Earnings management occurs when managers use a judgment in financial reporting and is structuring transactions to alter financial reports to either mislead some stakeholders about the underlying economic performance of the company



or to influence contractual outcomes that depend on reported accounting numbers.

Similarly, another definition of earnings management says earnings management is



a gray area where the accounting is being perverted, where managers are cutting corners and where earnings reports reflect the desires of management rather than the underlying financial performance of the company.

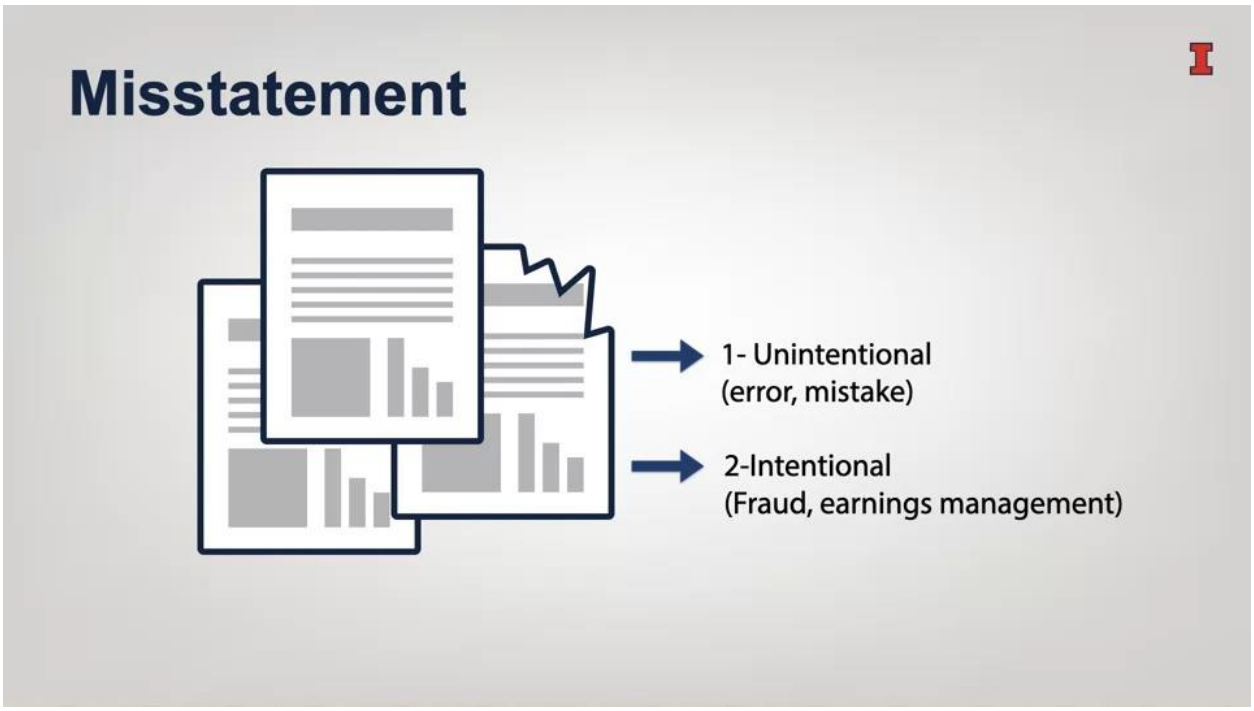
Although it's hard to know for sure, accounting researchers expect that fraud and financial reporting are rather widespread. [inaudible] in an accounting research study in 2018 estimated that the fraction of CEOs that misstate earnings at least once is about 60 percent. In a 2013 study, Dyke Morrison's and [inaudible] estimated that the probability of a company engaging in fraud in any given year is 14.5 percent. Graham Harvey, Raj Cupola in 2005, surveyed hundreds of Chief Financial Officers or CFOs and found that a large majority admitted to managing earnings even at the expense of long-term firm value. These managers admitted to sacrificing long-term value to improve financial reporting perceptions. They know that several Chief Financial Officers state that, "You have to start with the premise that every company manages earnings."

It's also clear that fraud in earnings management is costly in the long run to investors, creditors, bankers, and other users of the financial statements, including the companies that commit the fraud themselves. Dyke Morrison and Galus in 2010, study 216 companies with corporate frauds between 1996 and 2004. They found that the final settlements or punishments for the companies reflect a median fraud punishment of $34 million with an average of $198 million.

Lesson 2-3: The Fraud Triangle and Employee Fraud

[Music]

The job of a financial auditor is to make sure that the financial statements presented and recorded by a company, their audit client, follow the rules for generally accepted accounting principles.



This means that the auditor is looking for mistakes which are usually called misstatements. These mistakes can be both intentional and unintentional. This lesson focuses on intentional misreporting.

Of course managers and upper level management are not the only ones who commit fraud.

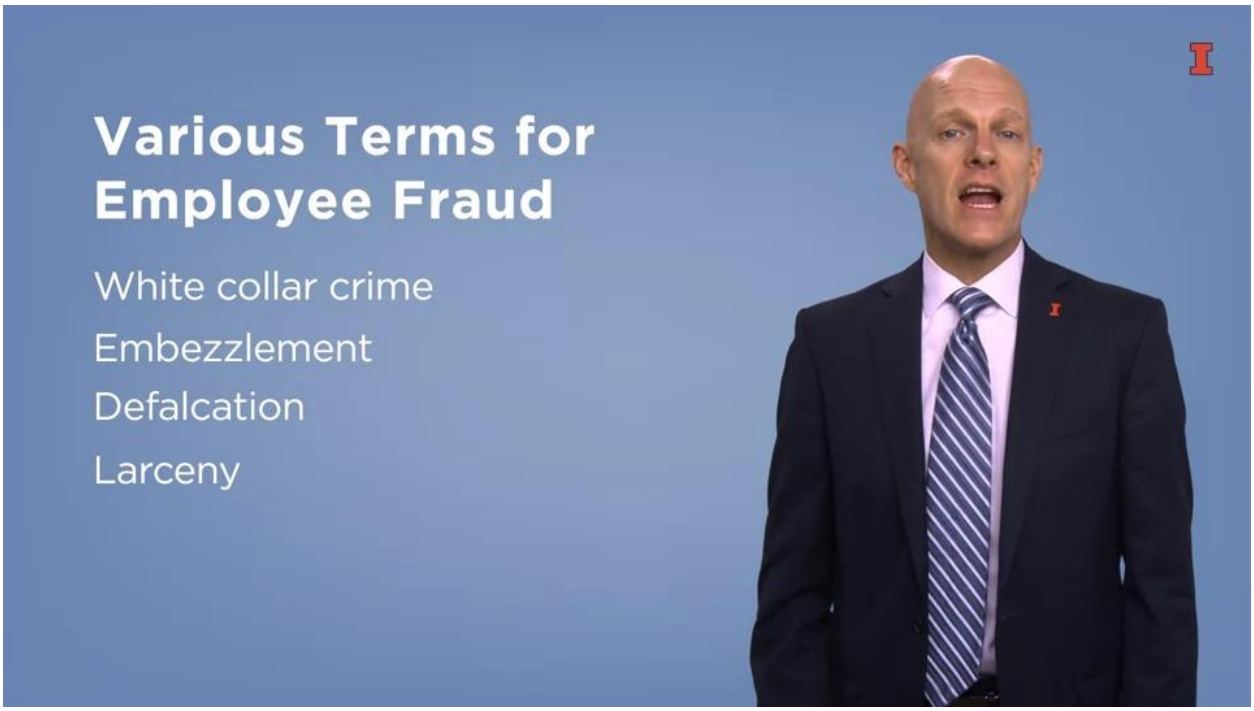
Middle managers and everyday employees also have incentives to cheat. This cheating may lead the financial reports to be misstated

and thus has to be a concern for the financial order. Furthermore, this type of fraud let's call it employee fraud is also significantly costly. Randy Wilson who is a partner and the national director of fraud services at RGL forensics claims that fraud can't be prevented.

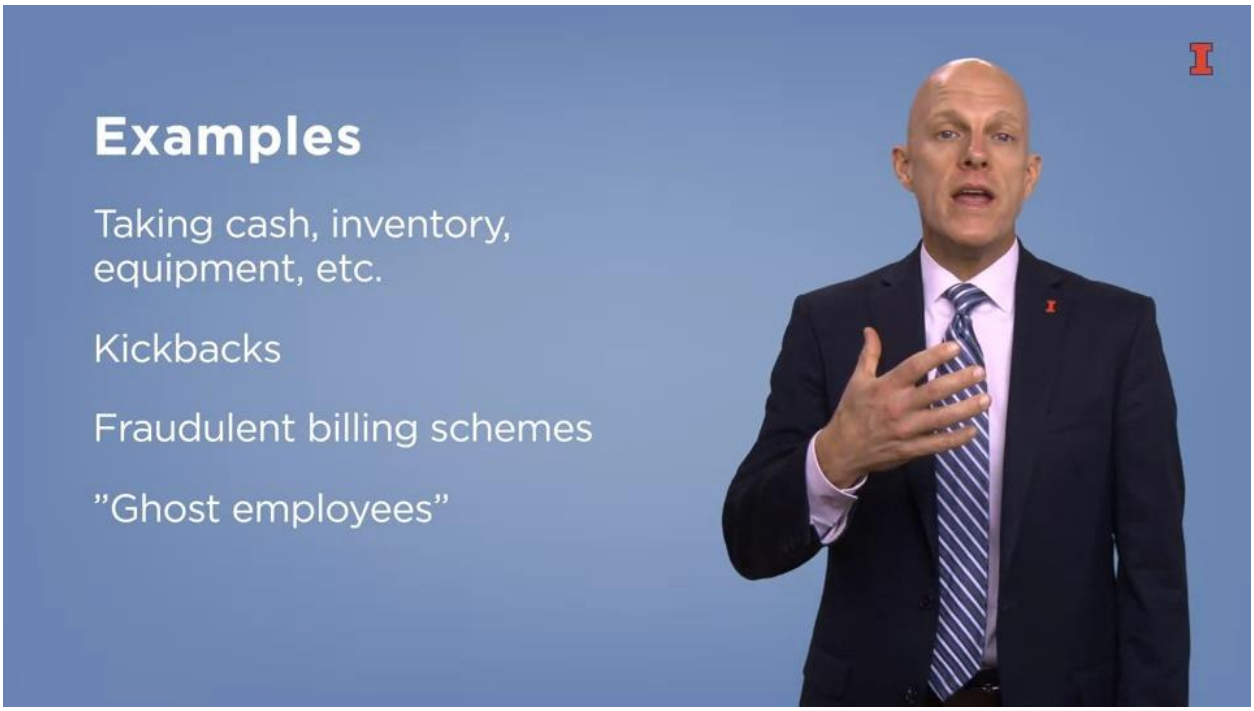
Other statistics show that 5% of business revenue worldwide, or 35 trillion is stolen through fraud every year. That only 49% of victims recover any losses. That one in seven employees take confidential information to a new job.

And that 31% said they would deliberately steal and share sensitive information if they were fired.

Employee fraud can have many different terms. One overarching term is white collar crime.



White collar crime is generally defined as fraud perpetrated by people who work in offices and steal with a pencil or with a computer terminal. Employee fraud can variously be called an embezzlement, defalcation or larceny. Embezzlement is a type of fraud involving employees wrongfully taking money or property entrusted to their care, custody and control. This is often accompanied by false accounting entries and other forms of lying and cover up. Defalcation is another name for employee fraud and embezzlement. Larceny is a simple theft of an employer's property that is entrusted to an employee's care custody or control.



Some examples of defalcation frauds include theft or misuse of corporate assets for the benefit of employees such as taking cash, inventory or equipment etc. Another one is kickbacks. These are ones special unapproved deals are given to suppliers and then those suppliers pay some of the extra money they make back to the employee that helped them get that good deal. Next fraudulent billing schemes based on the submission of fictitious supplier invoices might go directly to the employee instead of to the company. Another example is ghost employees. These are fake employees that are kept on the payroll after they retire or resign. The fraudulent employee can then fix things so that they get the ghost employees salary.

Employee fraud and upper management fraud is really hard to detect. This is why analytics are such a useful tool for auditors. Researchers have identified three factors that when present together lead to a higher probability that fraud will be committed.

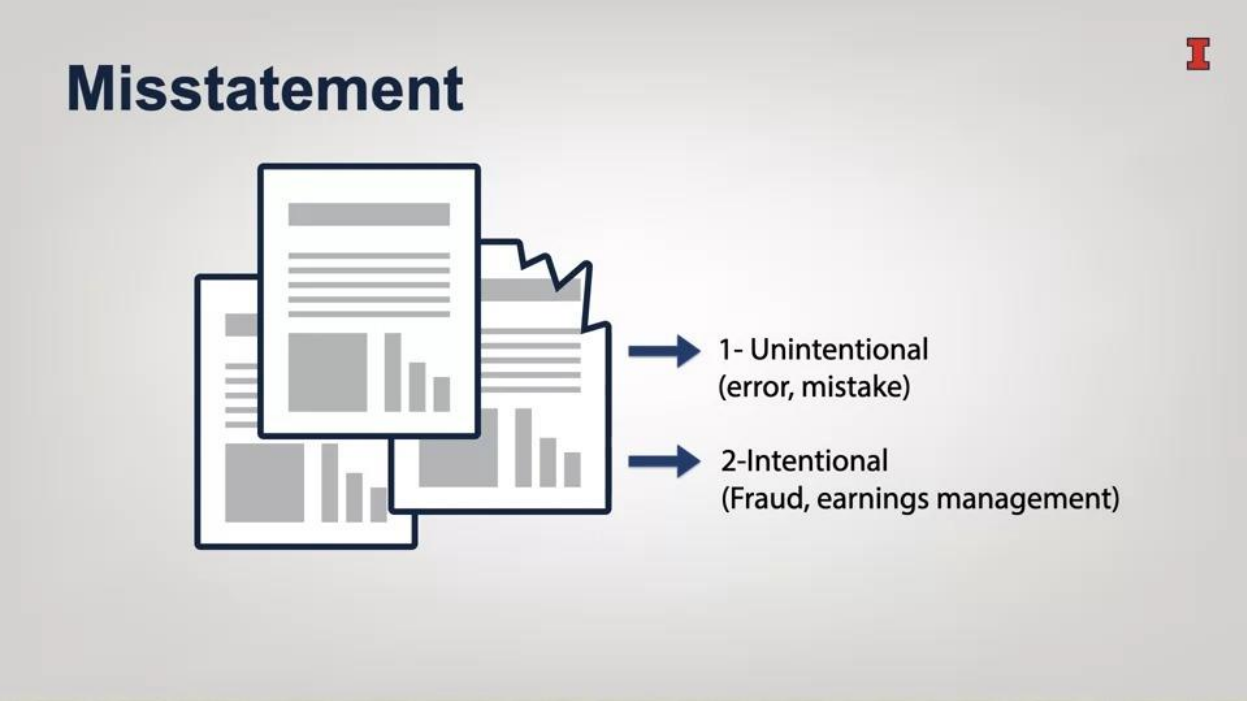
These three factors are jointly called the fraud triangle. The fraud triangle applies equally to upper management fraud and to lower level employee fraud.

The first factor is the motivation to commit the fraud. The incentive to commit the fraud or the pressure to commit the fraud. There are many motivations that employee might have to commit fraud, such as the desire to look good, desire to make more money or some internal pressures such as extensive or overwhelming medical bills. The second factor is the opportunity to commit from. This factor indicates that the employee has for whatever reason, the opportunity to commit fraud. This can happen when the company has improper passwords or other controls. It happens when one employee, for example, is in charge of many different aspects of the financial system, such as controlling cash and recording the receipt of cash. It also might happen when a trusted employee is allowed to be in charge of many different parts of the financial system. Finally, the third factor is rationalization or justification. Rationalization is when the employee has a way to justify the fraud. For example, perhaps they feel slided by the company. They feel that dedicated service over many years entitles them to something more. Maybe they justify that they'll commit the fraud, but only this one time or they tell themselves that the fraud is small and won't really hurt anyone else. The concept of the fraud triangle is important because fraud is hard to detect. Furthermore, white collar criminals are not generally like the stereotypical criminals we think of. They're just everyday people like you and me who are in circumstances that provide them the motivation for, opportunity for and rationalization of committing fraud. That is most people don't wake up one morning and decide, hey, I'm going to go commit a multi million dollar fraud today. Rather over time, a person rationalizes over and over and then slides down what is often called a slippery slope until they're committing a large fraud. Thus, an ardor should pay attention to when these three fraud triangle components are present, because that's when the fraud is most likely to occur.

Lesson 2-4: Need for Controls, Definition of Controls, Introduction of Internal Audit

[Music]

The job of a financial auditor is to make sure that the financial statements presented and reported by a company; their audit client, follow the rules of generally accepted accounting principles.



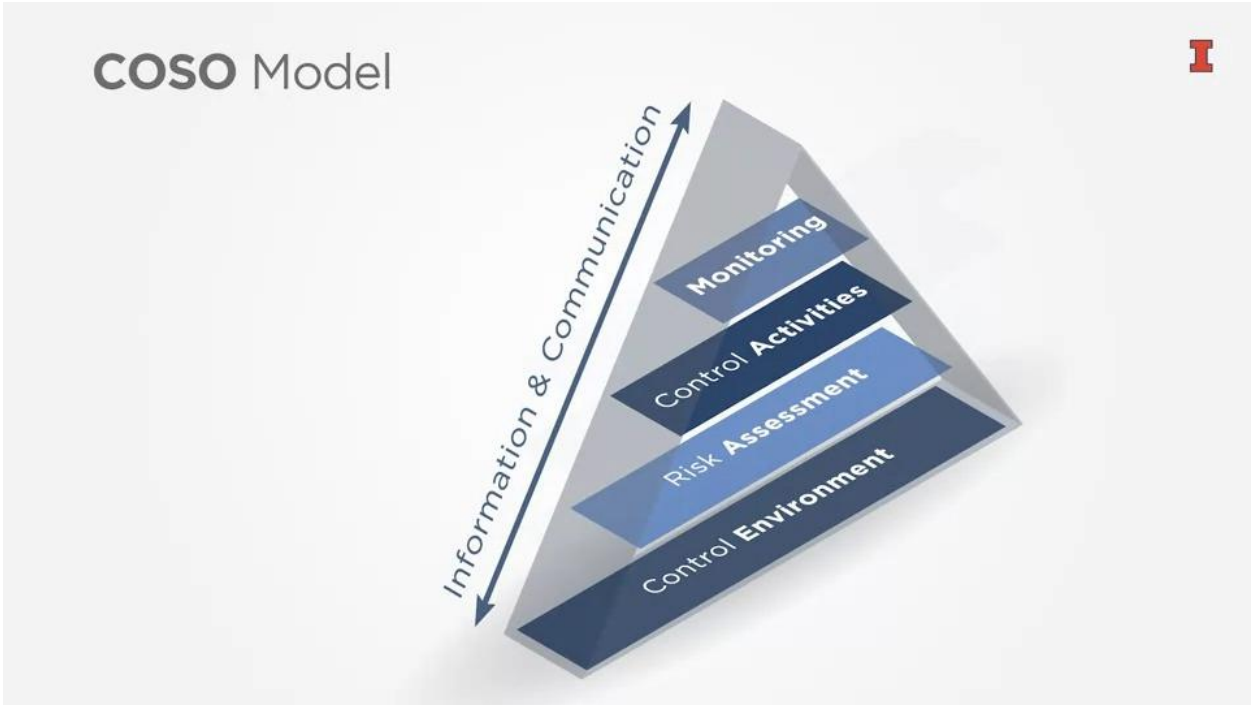
This means that the auditor is looking for mistakes. These mistakes can be either intentional or unintentional.

The majority of misstatements that an auditor finds are unintentional misstatements. That is, they are due to inappropriate attention, lack of efforts, lack of training, or lack of sophistication. Of course, knowing all the accounting rules is not easy and mistakes happen all the time. Furthermore people are not perfect and they make mistakes, thus the auditor must be ready for and aware of any misstatement whether it's from fraud or from unintentional error.

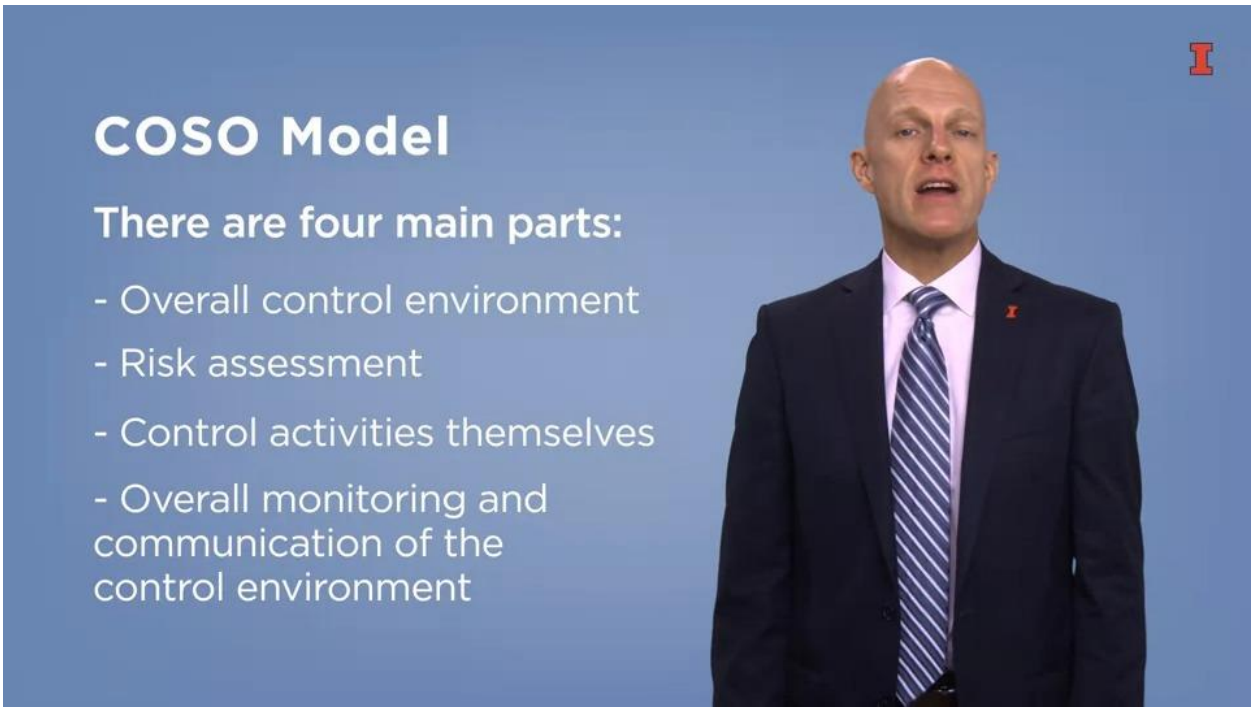
Of course, the management of the company knows its employees are not perfect, and they also know that fraud might occur. Finally, they know that it's their responsibility to ensure that the financial statements are free of error at least to a reasonable level, thus management puts in place internal controls that are structured and created to limit misstatement, fraud, theft, etc. The auditor is employed to test and evaluate these controls in addition to their job of ensuring that the financial statements are not misstated.

The basic idea behind internal controls is that it's far easier to prevent a problem from happening than to fix the problem after it's already happened. You've undoubtedly seen internal controls at work or just going about your everyday life. Think about the locks that you see on the bank, the security guards at a retail store, the need for the manager to come over the grocery store checkout when you need to change the numbers of the items that were scanned. All of these controls are put in place to minimize the mistakes and frauds that often happen. In a way,

the internal controls act as a shield or a roof to protect the company from risk that would rain down upon the company if left unprotected.



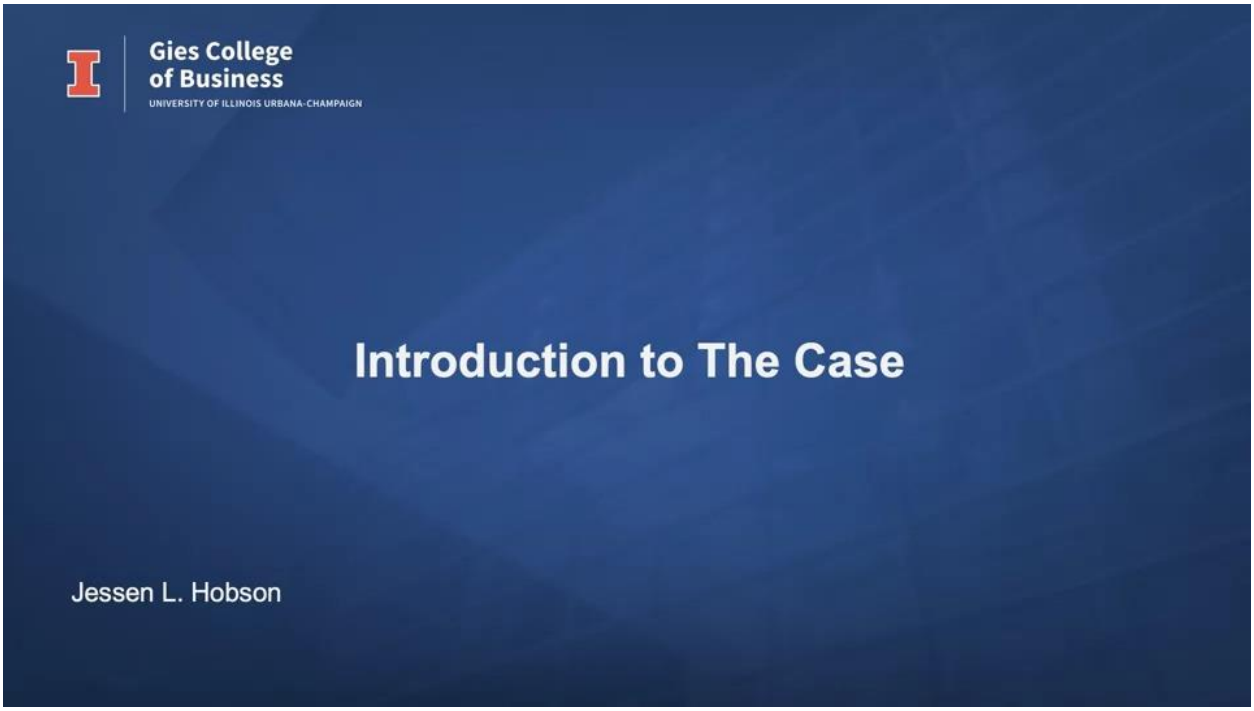
The components of internal controls over financial reporting created by the Committee of Sponsoring Organizations of the National Commission of Fraudulent Financial Reporting comprises of four main parts. First, there's the overall control environment. Second, there's risk assessment. Third, there's the control activities themselves. Finally, there's overall monitoring and communication of the control environment.



The control procedures comprise the heart or the guts of the whole control environment. These controls are the ones we discussed before such as physical controls over the security of assets, physical safeguards and access restriction, proper authorization and approvals. They also involve separation of duties where one individual is not responsible for too many duties. Other control activities consists of company performance reviews, information processing controls such as: passwords, firewalls, and two-factor authorization, and reviews, and reconciliations.

Once controls are in place and communicate to the employees of the company, management often wants someone internal to the company to check on these controls and to make sure they're working properly. These individuals are called internal auditors. Whereas the financial auditors we've previously been talking about are external to you or outside the company, the internal auditors are inside the company hired by the company to work for the company. Internal auditors are a critical part of the control system of a company. They do a wide variety of things, many of which are similar to what the external auditor does. They test the company's financial statements to make sure they follow the rules. They test the controls the company has in place to make sure they're designed to work properly. The external auditor as part of their job of auditing the company will want to make sure that the internal audit function is well-staffed and working properly.

Lesson 2-5: Introduction to The Case

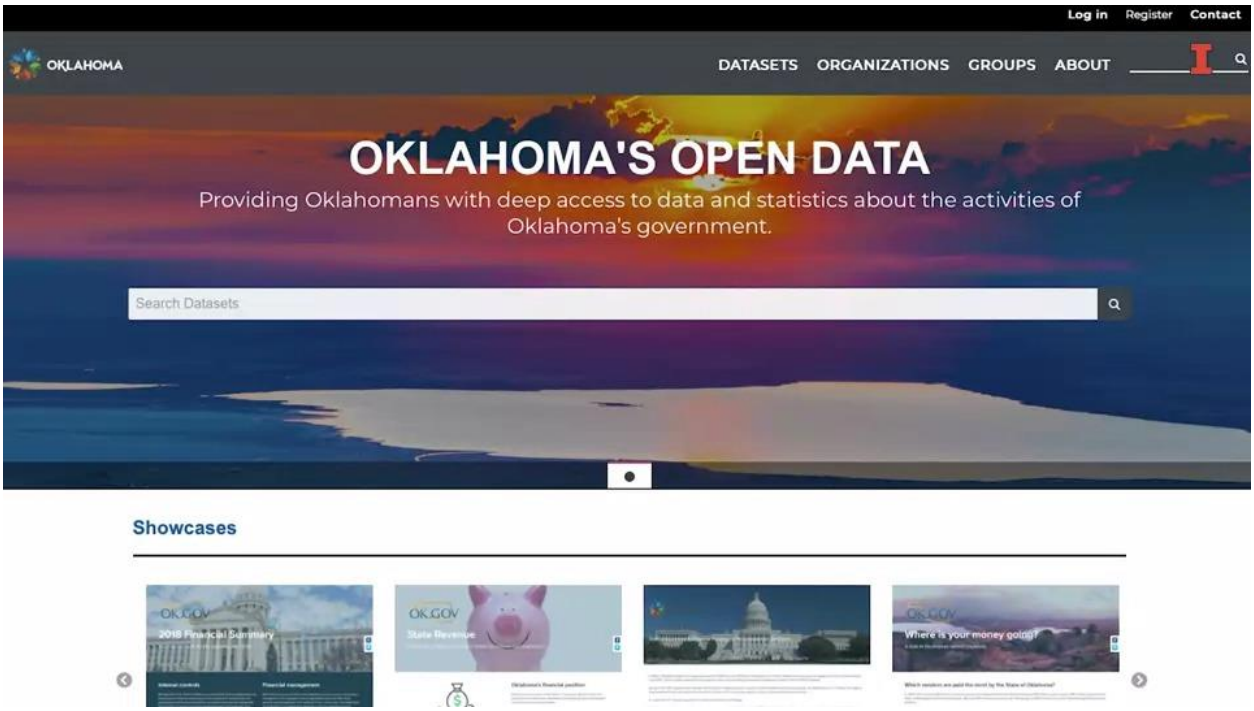


[Music]

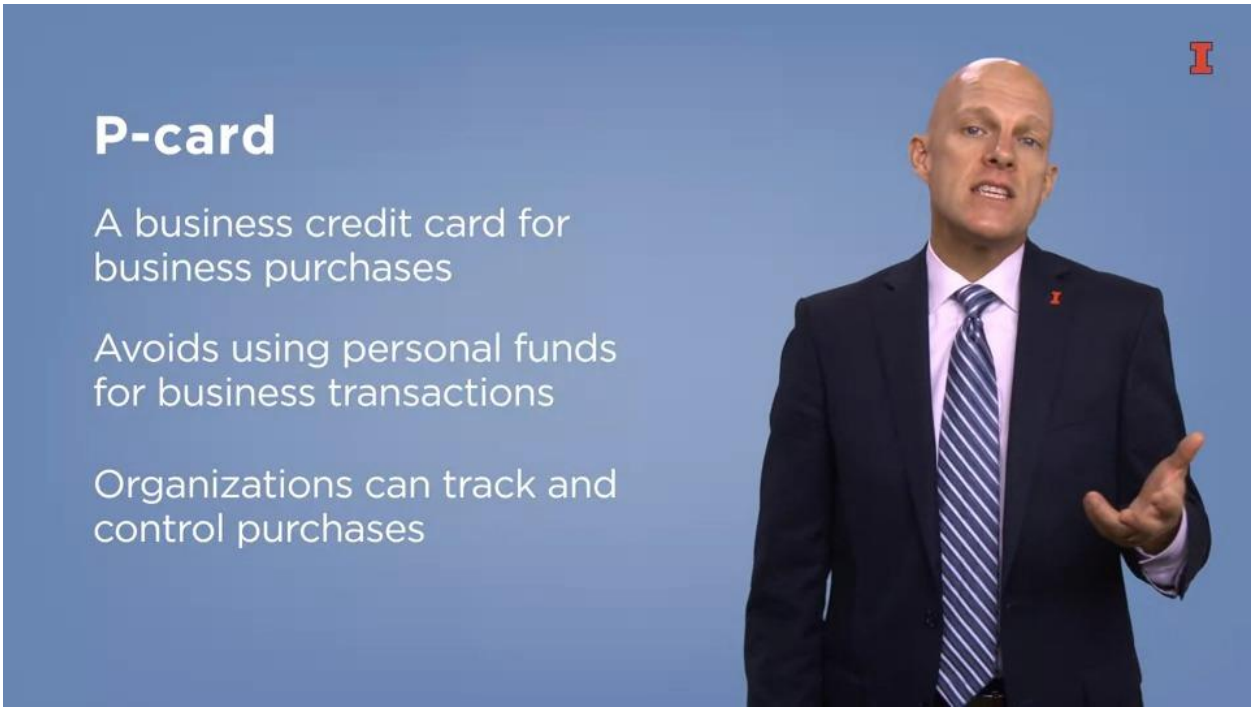
In this lesson, I'll introduce a case that we'll use to examine the audit of controls over credit card transactions.

This case is based upon a case from the Ernst and Young Academic Research Center, also known as EYARC. The case is called P-card. The use of the case material complies with the copyright therein. Specifically, the EYARC copyright prohibits posting case solutions on any publicly accessible platform. Thus, before using these materials, I modified all requirements and their respective materials. That is, no solutions to the original case are provided here.

The case uses real data from Oklahoma State University, also called OSU. Now Oklahoma State University is a public university funded and overseen by the State of Oklahoma. Now the State of Oklahoma like many other municipalities in governments in the United States makes a considerable amount of data publicly available. For example, the State of Oklahoma makes the following datasets available,

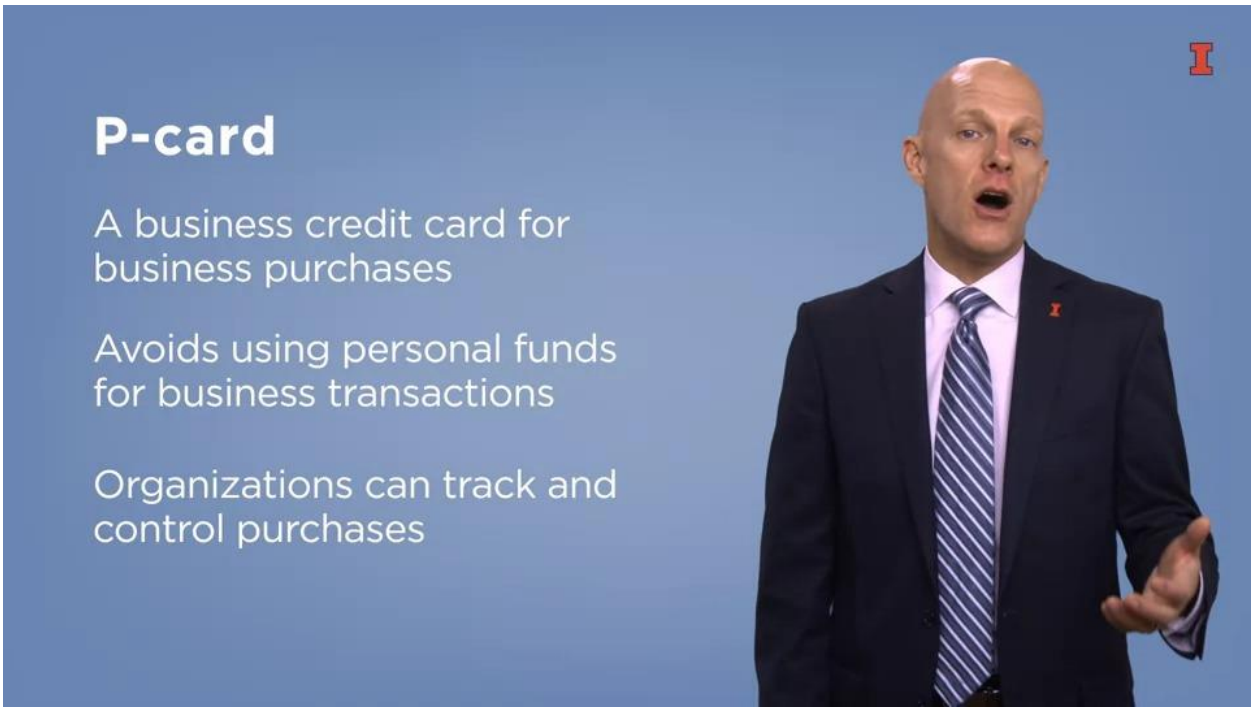


State of Oklahoma payroll, used motor vehicle and parts commission dealer license search, groundwater safety, and government purchases.



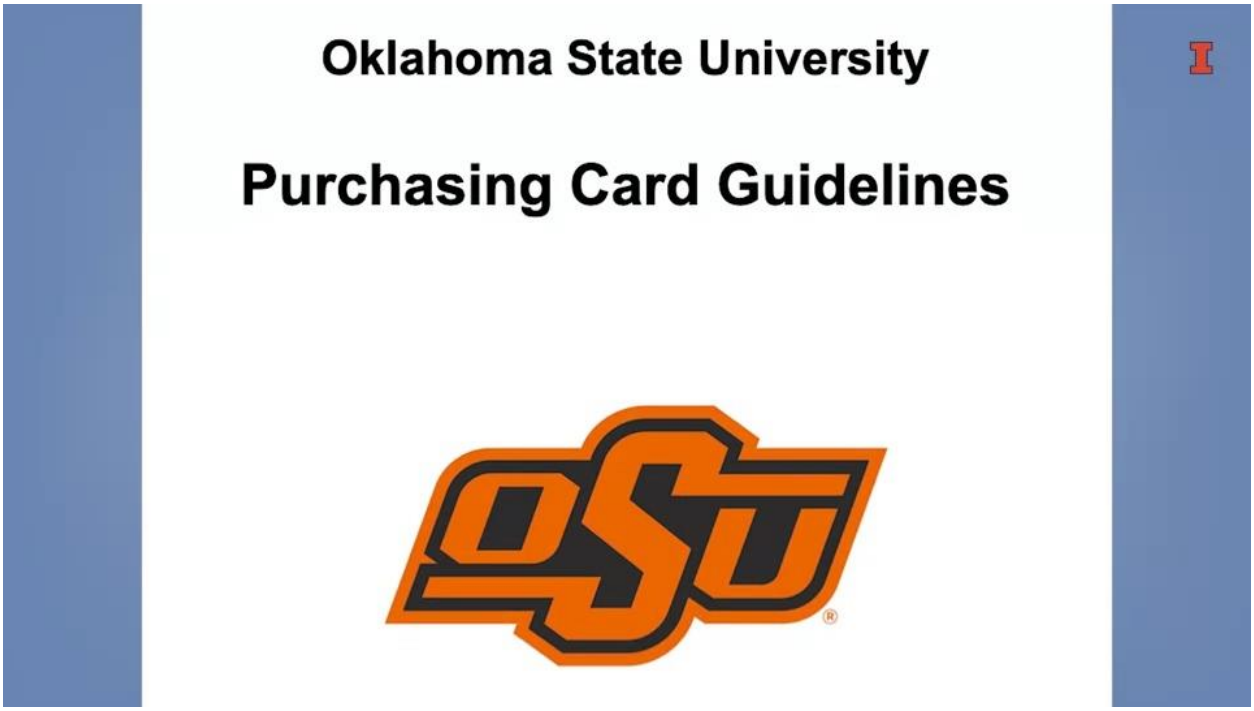
Now a cool data set that we'll use for this case is from the State of Oklahoma Payment Card Purchases. Payment cards or P-cards are business credit cards that some employees are permitted to use to purchase necessary goods and services for work.

If employees agree to certain rules, they can then use the P-card to make appropriate business purchases



rather than having to use their own credit card. This allows the employee to avoid spending personal funds and then later seeking for reimbursement. It also provides the business with greater control, because a business can institute internal controls to limit types of purchases, and avoid inefficient and fraudulent transactions. Organizations can also track spending using detailed P-card records provided by the credit card companies. The Oklahoma state government does not want its employees to use P-cards for just anything. Now P-cards are only allowed for authorized business purchases. In order to make sure this is the case, the organization then authorizes the use of P-cards has rules or controls about what a user of a P-card is allowed to spend money on, and how much they're allowed to spend.

Our job in this case, is to act as an auditor engaged to audit the P-card transactions to ensure that those using the P-cards are following the stated controls.



Specifically, this case tosses you into the role of an internal auditor for Oklahoma State University. You're asked to perform an audit of the purchasing cards, the P-cards that are used on campus.

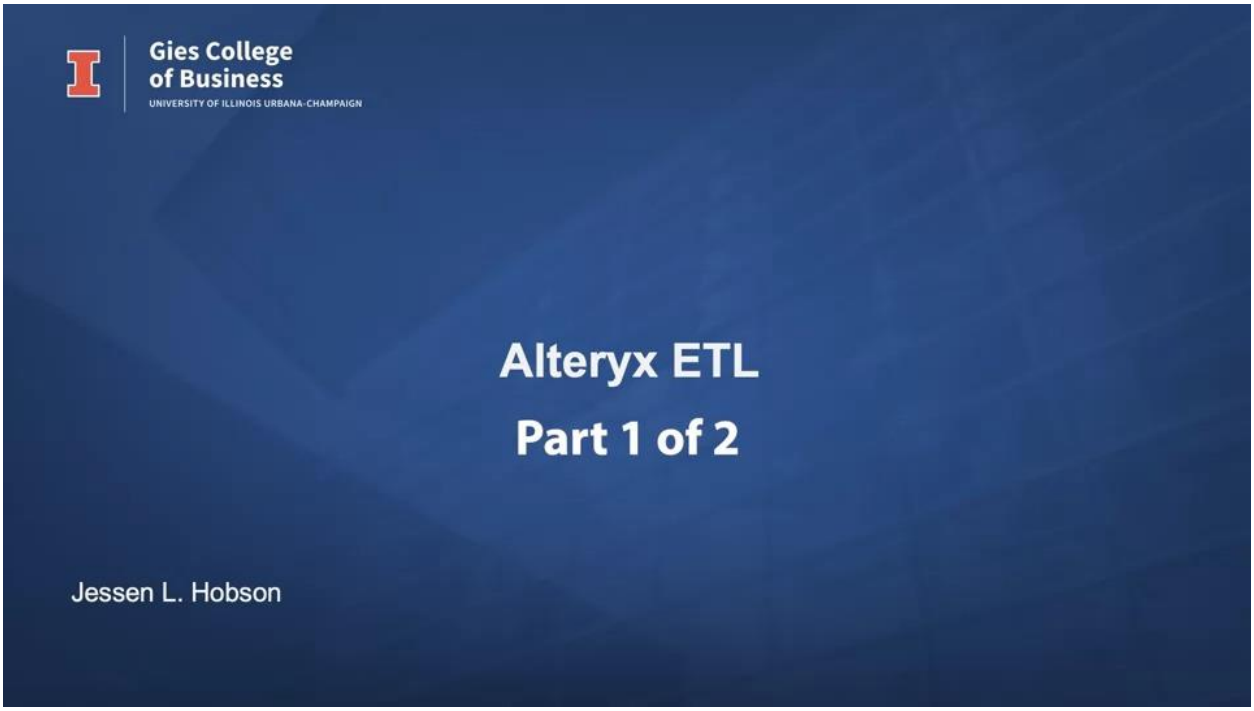
There are three specific controls that you're asked to check. First, total purchases for the year are $50,000 or less. That is, an employee is not allowed to make purchases totaling more than $50,000. Second, employees are not allowed to make purchases totaling more than $10,000 a month. Finally, individual transactions are not allowed to exceed $5,000 each. Let's pause here and examine these controls and think a little bit more about each one. Specifically, let's first think why they're important and why OSU might have implemented them. That is, what are the risks that these controls are mitigating?

Let's think first about this from a high level. First, what's the overall risk when employees are using purchase cards? The main risk, broadly, is that employees are making purchases that are not authorized. That is, employees are making purchases of items they're not allowed to purchase. There are multiple ways to control this. One way would be to restrict which organizations the employee is able to make purchases from. While this would be a useful control, it would probably be hard to regulate since there's probably many organizations from which the university may need to make purchases. Another control might be to restrict which items the employee purchases. Again, this would be a very effective control, but would be hard to implement since it's difficult to regulate what items are purchased, and even to view what items are purchased in each P-card transaction. Thus, a simpler control may be to limit the amount of purchases that are made by each employee such as in the controls discussed above. This is not a particularly precise control since it doesn't get directly at unauthorized purchases, but if the limits for how much someone can purchase are set at reasonable levels, it will prohibit high-value purchases that are likely not for business purposes. Another risk with any transaction is that it's not properly approved. Approvals might be in place for transactions to ensure that they follow state guidelines, do not exceed budgeted amounts, do not require a purchase order, and do not require a formal bidding process through a formal request for a proposal or RFP. Spending limits are excellent controls to limit these types of risks.

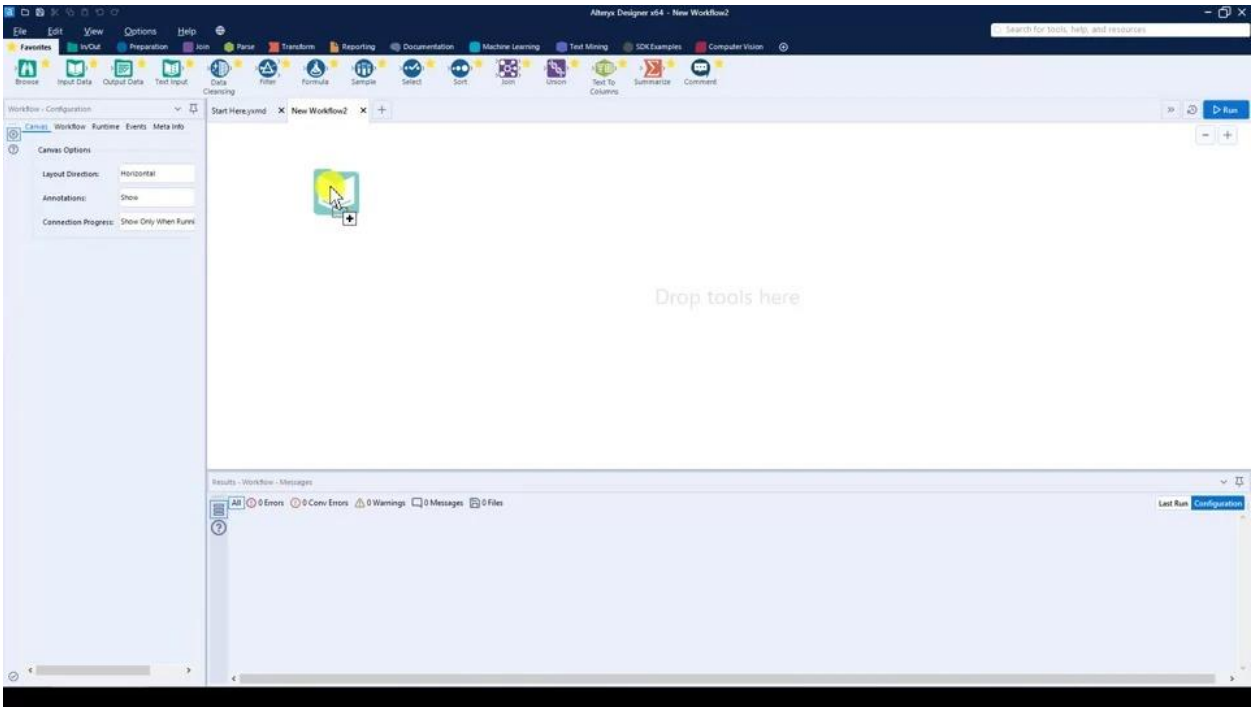
Second, let's think about how we can implement these controls. In general, there are two ways to implement controls, in a preventative manner or in a detective or oversight manner. In a preventative setting, the controls are set up such that they do not allow an employee to do a certain thing. For example, in the case of the third control above in which an individual transaction is not allowed to be greater than $5,000, a preventative control would make it impossible for employees to conduct that transaction that's greater than $5,000. Now a second way to implement these controls is as detective or monitoring controls. Such that the employee is not prevented from doing something, but rather a log or a record is kept, and after the fact, the record can be examined and reviewed to see whether or not an employee has made a purchase for over $5,000, for example. While preventative controls are generally superior to detective controls, they're often much harder to implement. For example, it's not clear how OSU would prevent employees from making transactions of greater than $5,000. Furthermore, there might be situations in which an employee should be allowed to make a purchase that's greater than $5,000. Thus, in the case of OSU, these three controls are implemented as detective controls. Your job then as the auditor is to go into the data to examine whether or not these controls have been violated. Finally, let's think about how employees might circumvent, subvert, or get around these controls. For example, an employee might try to get around the control that limits individual purchases to less than $5,000 by splitting a larger transaction into multiple smaller transactions. While we don't examine this possibility in this case, it's important for the auditor to keep an open mind and to think skeptically about how an employee might circumvent the system.

Our approach in this case will involve three parts. First, we'll load, view, transform, and clean the data to make sure it's ready for use. This process is often called ETL for extracting, transforming, and loading data, data merging, data prep, among many other names. Data analysts frequently claim that 80 percent of their jobs are getting and cleaning data. Thus, it's important for us to practice these skills. Second, once the data is loaded and cleaned, we'll explore it a bit. This process is often called EDA or exploratory data analysis. In this case, we'll have three specific controls that we're testing. Nevertheless, we should still seek to understand the data. Doing so might reveal additional problems or issues that we need to consider in our control testing or that we want to follow up with for additional testing or inquiry with the audit client that we're working with. Finally, we'll check all of the data to examine whether any of the three controls have been violated. We'll then make a list of violations and follow up with management to see if they have explanations for these violations.

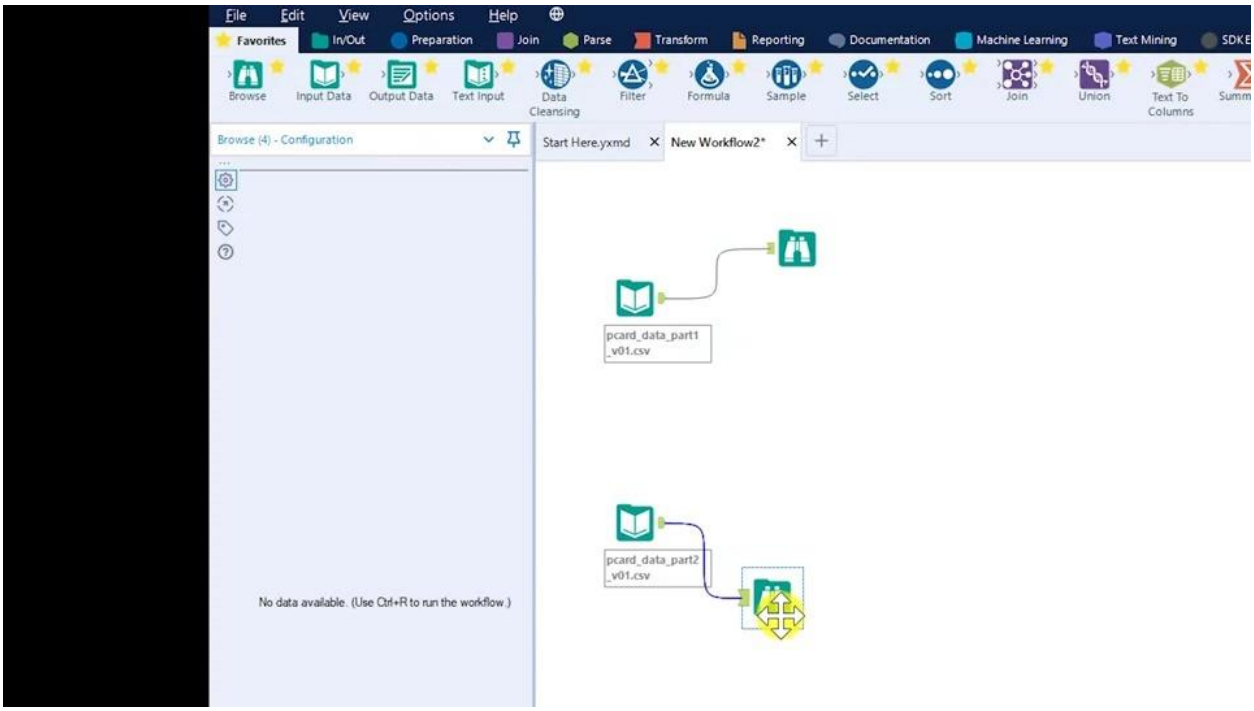
Lesson 2-6: Screen Shot Solutions in Alteryx - ETL Part 1



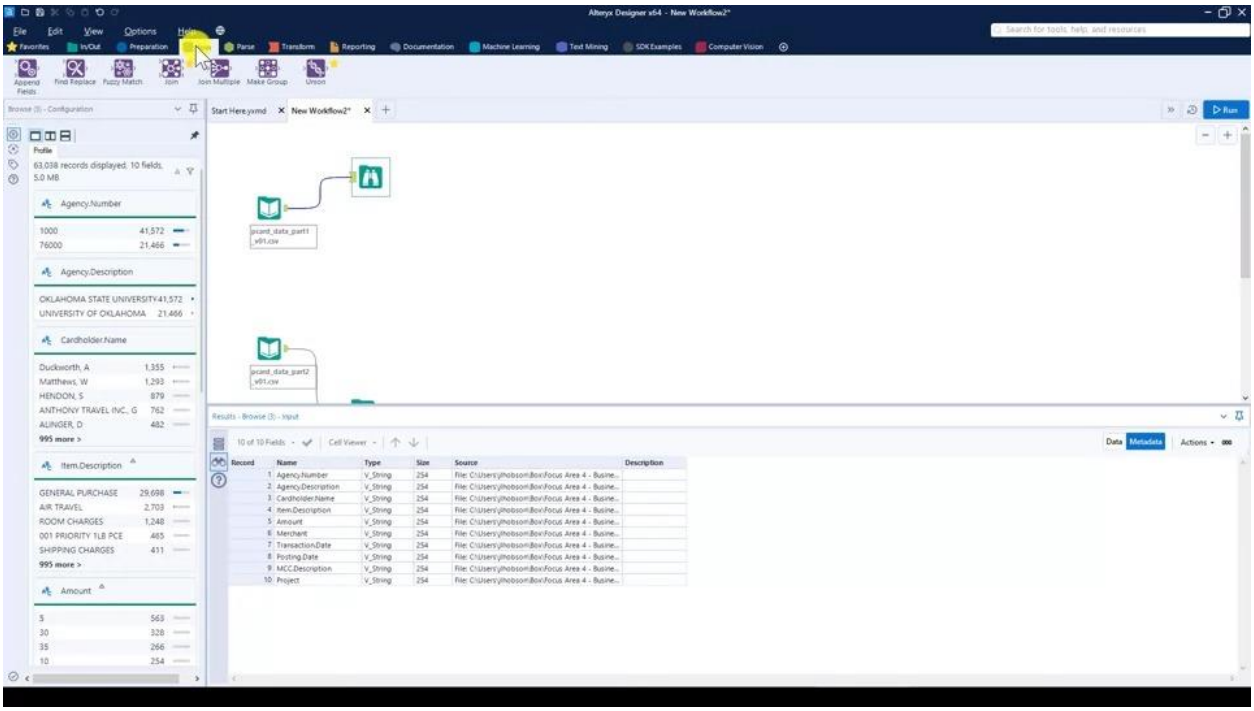
[Music]



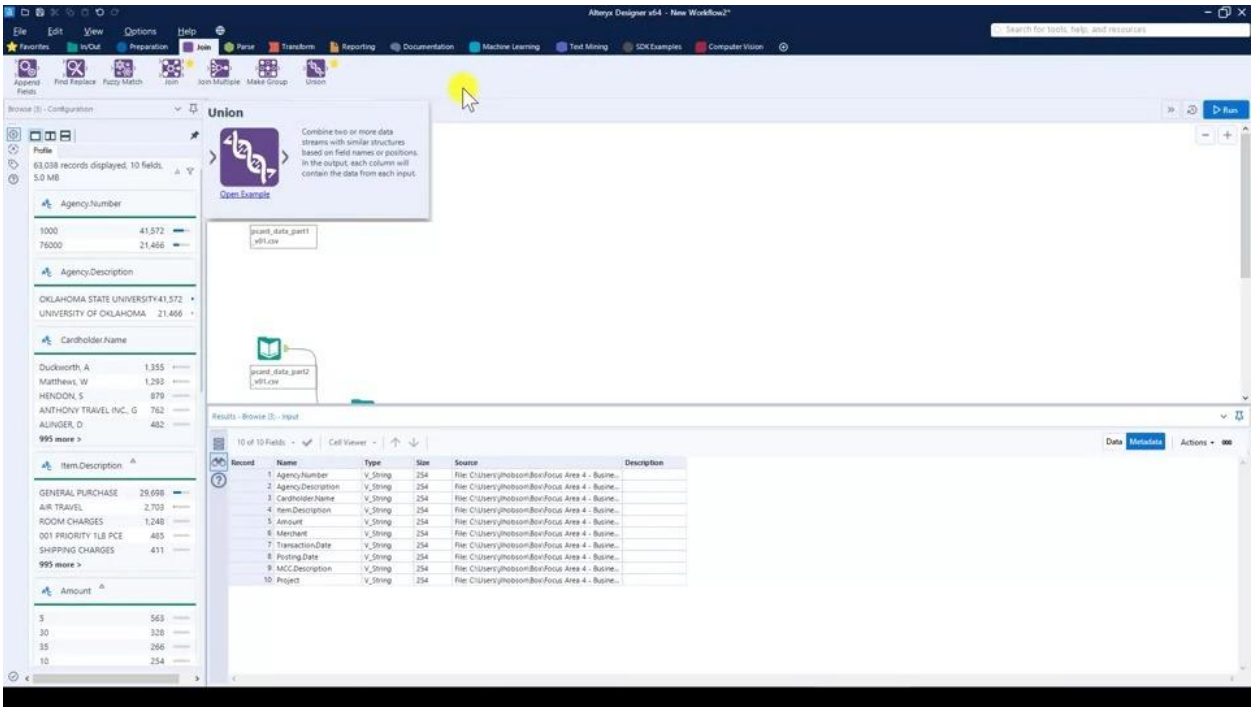
Let's start our case. The first thing we need to do is to load the data. The case includes two separate datasets, so let's load those into Alteryx. Then let's see what we can learn about them. We'll bring in a Browse tool from the Tool Palette to look at them and view them and see what we can learn. Let's bring both of those in, there's two separate ways to bring data in. First, I can just drag it in, so I'll go off-screen and go ahead and grab that first file. This is pcard_data_part1_v01 it's a CSV file.



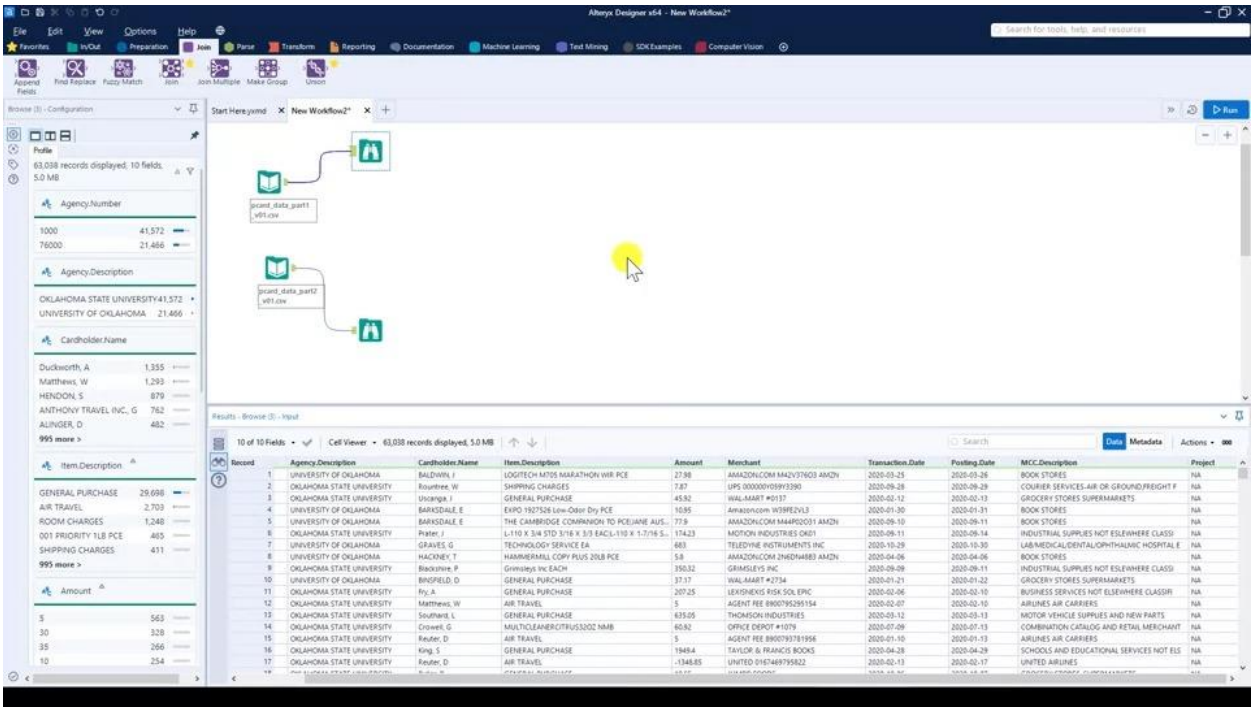
I'll just drop that in and you can see the names there and it's all ready to go. Secondly, I can go up to the Tool Palette and I can get the Input Data tool right here and pull that down, and then I can go over to the Configuration Panel over here and find the path that I want and pull that in. I have it just right here because I've pulled in earlier. I can do that. Once I click off of it, you'll see the names there and it's all ready to go. I have both of my two files that I need to do, and I'm going to go ahead next and I want to look at the data. There's at least two ways to do that as well. First, I'll go get the Browse tool from up in the Tool Palette and I'll drag that over. That will allow me to look at the spreadsheet once it comes in. I'll put it on both of these, I guess. I'll do that here.



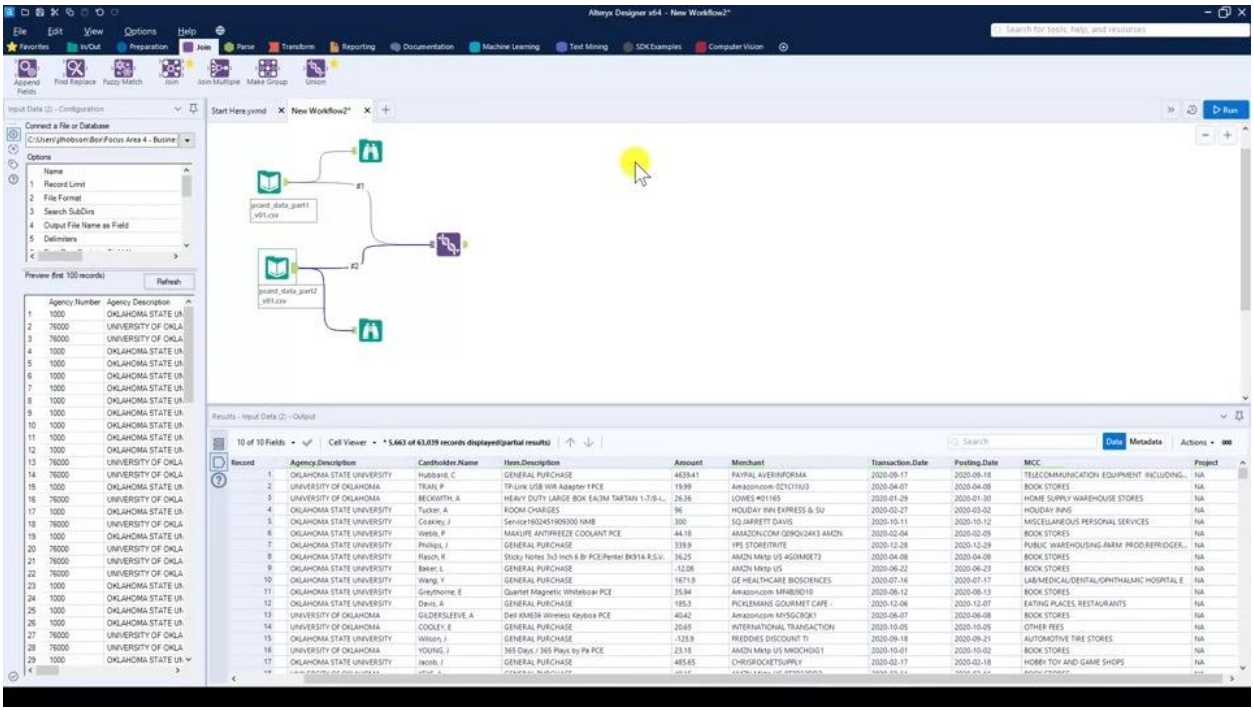
I'm going to run things and we'll show you the two ways to look at this. I run it everything looks good. I don't see any errors down in the results window. Let me look at my first Browse tool, and I've got a couple of things I can look at. On the left, I've got a couple of different ways here in the configuration window to look at the data, and so it's going to give me each of the columns here and give me this nice summary information about it so that's pretty cool. That's nice to use. The other thing I can do is I can click on the "Output Anchor" here so I can look down the results window, and this gives me every column. I can look here and I look through each column and it gives me a little bit of data there for each one so that's also useful. Of course, I can click up here at the "Messages" area of the results window to see that there are no errors. If I go back to the Output Anchor, I can see the number of rows that were brought in, and it only shows some of those rows just for size limitations. But I can also get Metadata over here, and that's also helpful. What I can see as it's brought in strings, it brings everything is in a string, so we'll have to deal with that here in a bit as well, of course. What we need to do next is we'd like to join these two different datasets together. There's multiple ways to join data together.



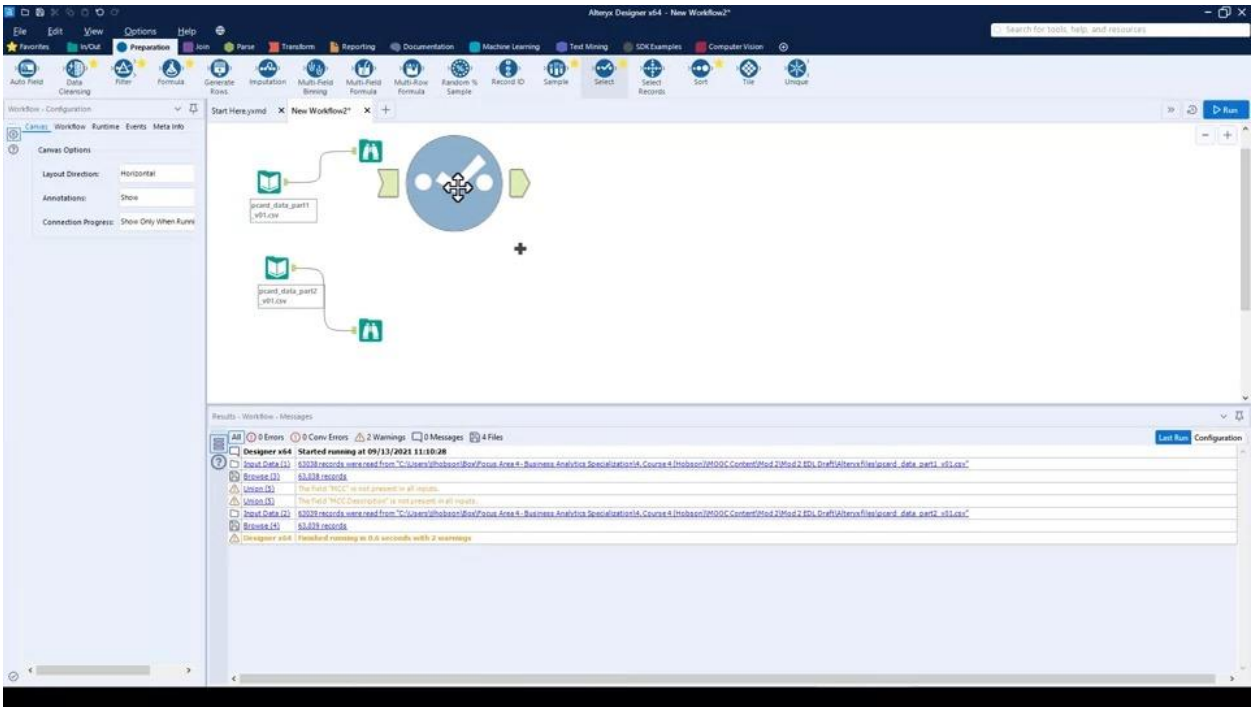
If we go to the Join area up here in the Tool Palette, we can see several, and let's go ahead and look at Join which sounds like exactly what we want to do. Because we're trying to join things together. This has combined two data streams based on common fields. Normally what we have here, when we use the Join tool, is we have two separate datasets that have one common column, a primary key that is the same again in both datasets, but then different columns in each dataset. We bring them together, merge or join is what we would call it in R or an SQL. Then if we look at Union, here's our other options. This combines two or more data streams with similar structures based on field names or positions. The Union then is just concatenating the data. Here we want to stack rows together one on top of each other, or stack columns one after the other. This is what we want and why?



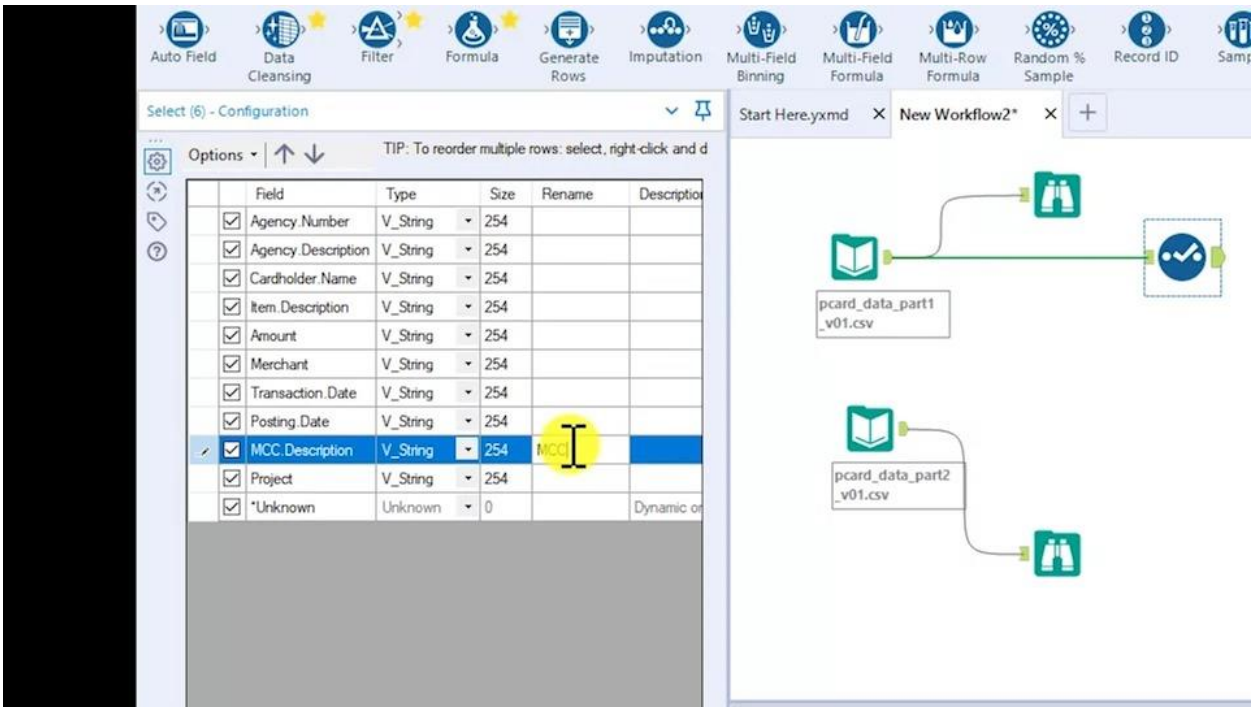
Well, if we look down in our data, I'm going to just pull this up, we can see that we have a case, here we go, where we have, if we click on "Data" in the results panel, we have 10 different columns, agency number, agency description. If we look at that and then we go up and look at the first one, we can see that we have the same columns. It looks like the data is just split in 1/2. What we want to do here is stack the rows on top of each other because they have the same column names and they'll just stack together, and that is the Union tool.



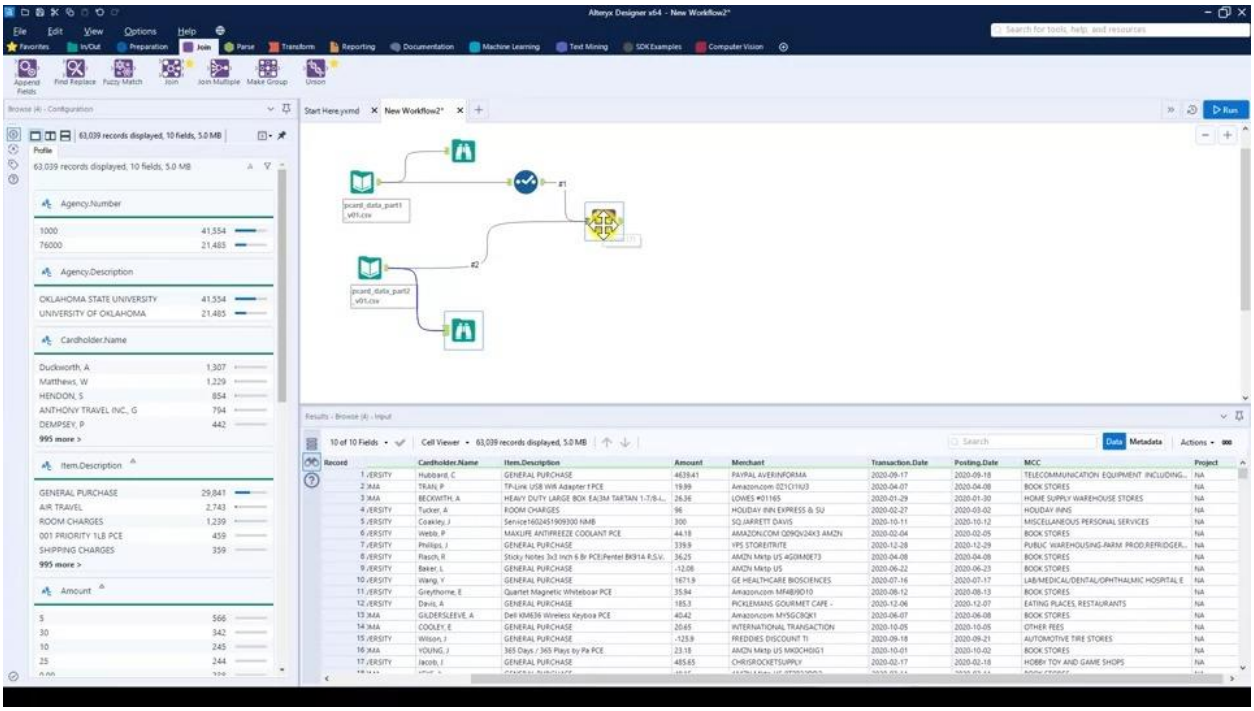
We bring the Union tool down and we're going to drag over a connection here and drag over a connection here, and just go ahead and hit "Run" and see what happens. That runs through. We looked down at the results window and we see no errors. But we do see this yellow caution. Says the field MCC is not present in all inputs, and likewise, field MCC.Description is not present in all inputs. We do have two warnings and that's problematic. Let's click on the "Results Anchor". Let's look through and sure enough we see that the first problem because we have 11 of 11 fields, and again, we want 10 fields. We're stacking together the rows from 10 columns. That's not what we get. We have a problem and as we scroll over, we see the nature of the problem as we look at each of these the combined results here, we see that we have MCC.Description, but then we have MCC. This was the warning we got before, and we can see that this column is null here. If we had all of the data, we would notice that MCC is null as we go further down. The MCC column is misnamed. It's named one thing in one of the datasets. In the 1st data set, it's called MCC description, in the 2nd dataset, it's called MCC. We just need to rename those to one common thing and we will be able to stack the data. The next thing for us to do is to fix that column. We want the columns to have the same name so that we can bring in these two datasets together. Probably the easiest thing for us is to go up to our first one, and then we can change that one to be, let's see, we have the MCC.Description column. Let's just change this to be MCC. Then we'll have MCC there, will have MCC here, and everything should stack together. We should have 10 of 10 fields and everything should work.



Let's delete this union here because that one didn't work. Then, we need to change the name from that first data set. Let's go up to the Tool Palette, and look under Preparation. It's not immediately clear which of these is going to be the best thing to use to change a name. One might be Data Cleansing, this is not going to work. It says perform basic data cleansing operations such, none of that says changing names. The right tool is the Select tool. Here if we read the Select tool, it says select, deselect, reorder, and rename fields. Perfect. That's what we want. Let's drag this guy in and we're going to put it here, I'm just dragging it over a little bit.

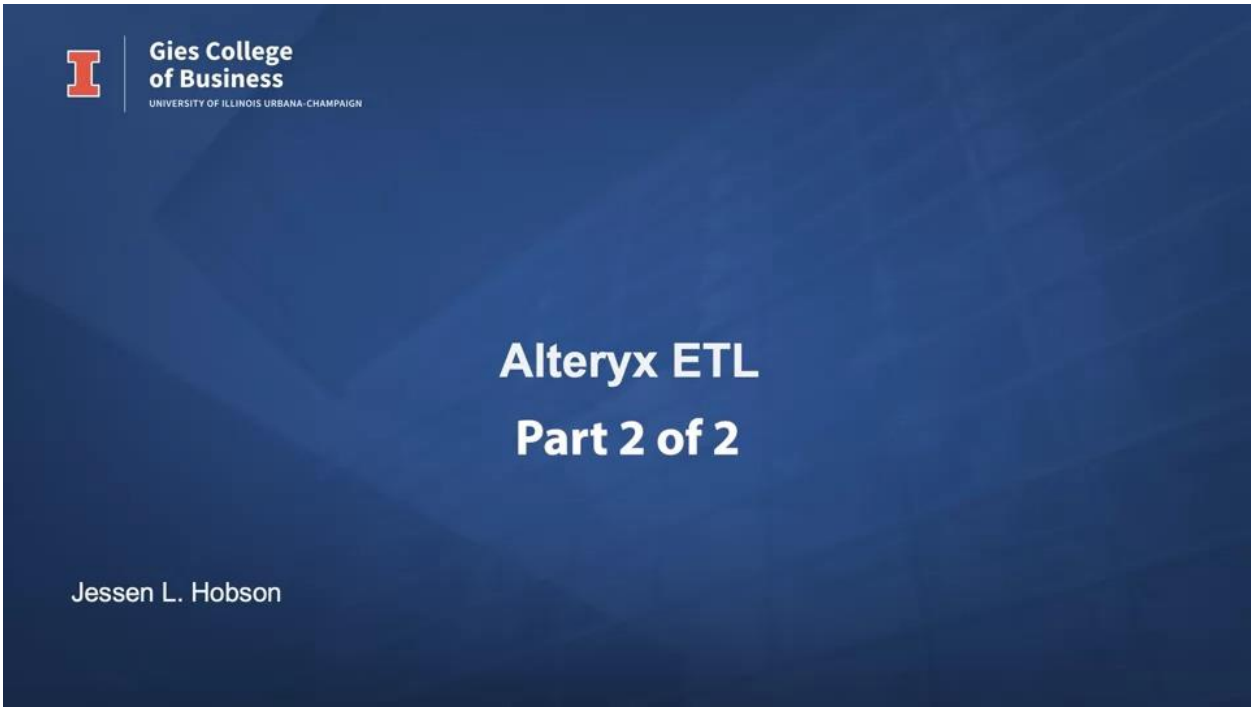


If we go over here to the configuration panel, we can see that there's this rename column. This couldn't be easier. All we have to do is go down here, find our MCC. Description and we're going to type in MCC, and we're all set and that should actually work.

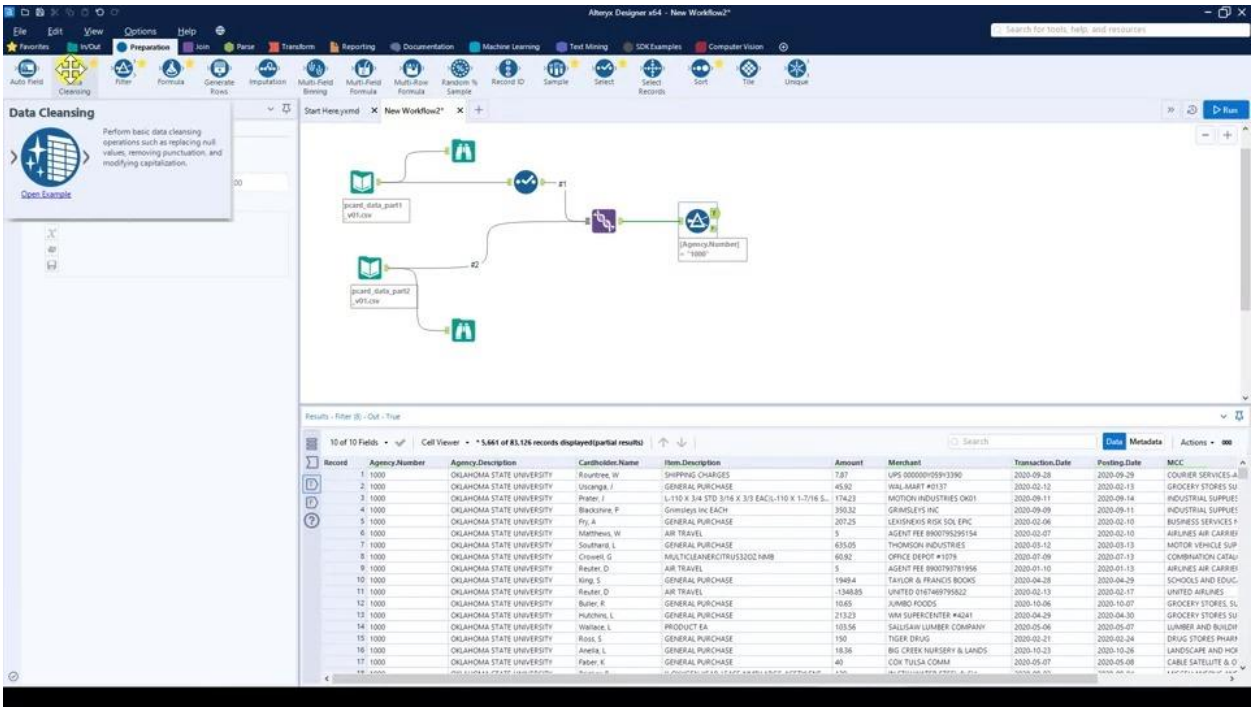


Let's go ahead and run that and see what happens. We've run it, we click here, we've got no errors or messages, and we can look at our Output Anchor. Clicking here the "Output Anchor" and now we have that first data set renamed MCC as we would expect. That's excellent. All set there. Let's go then to our Join and we'll pull down our "Union", and we want to union these two together, so we take that dataset, the first one cleaned at now, and we take our 2nd dataset and let's run that. Let's look at the output here. We click on our "Output" tab on both. We were looking at our Output tab we've got 10 of 10 fields, which is exactly what we want in columns. We've got our MCC column and no weird MCC.Description null or whatever. That looks great. We've got 126,077 records. Well, is that right? Let's see, we have 63,038 for our 1st data set, we have 63,039 for our 2nd data set, and if we do the math, that gives us 126,077. Everything great so far we've brought our two datasets together correctly.

Lesson 2-7: Screen Shot Solutions in Alteryx - ETL Part 2

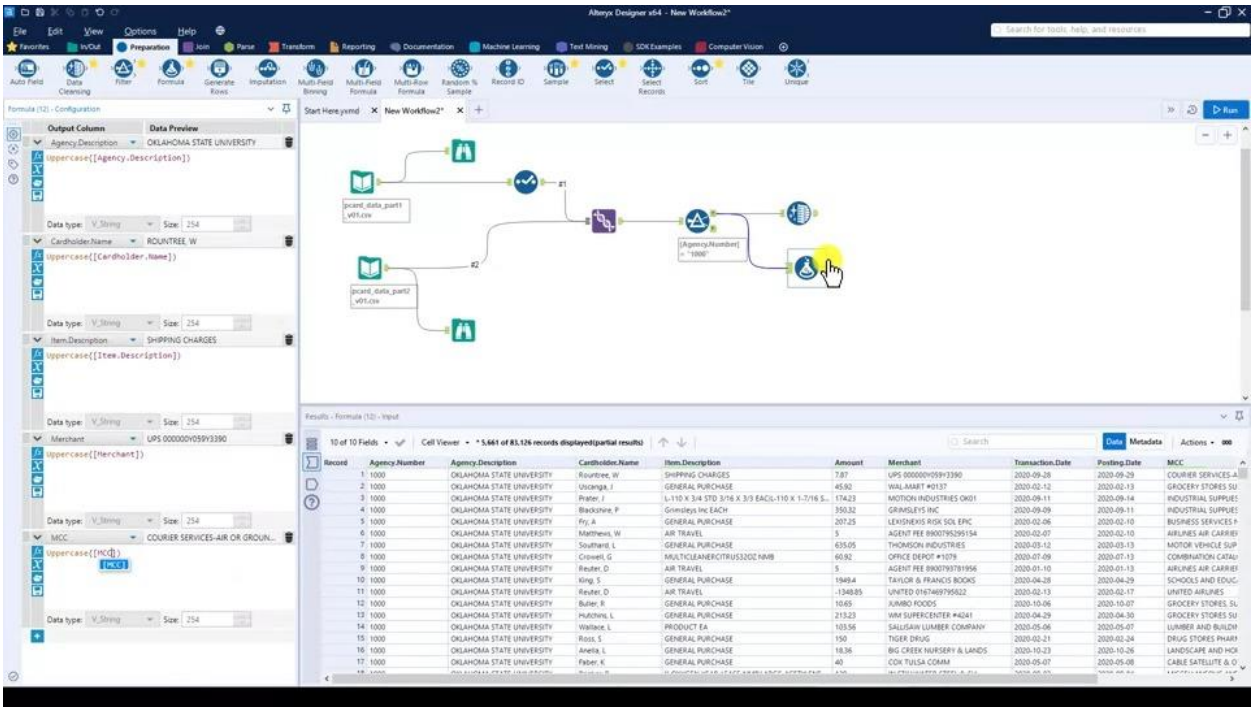


[Music]



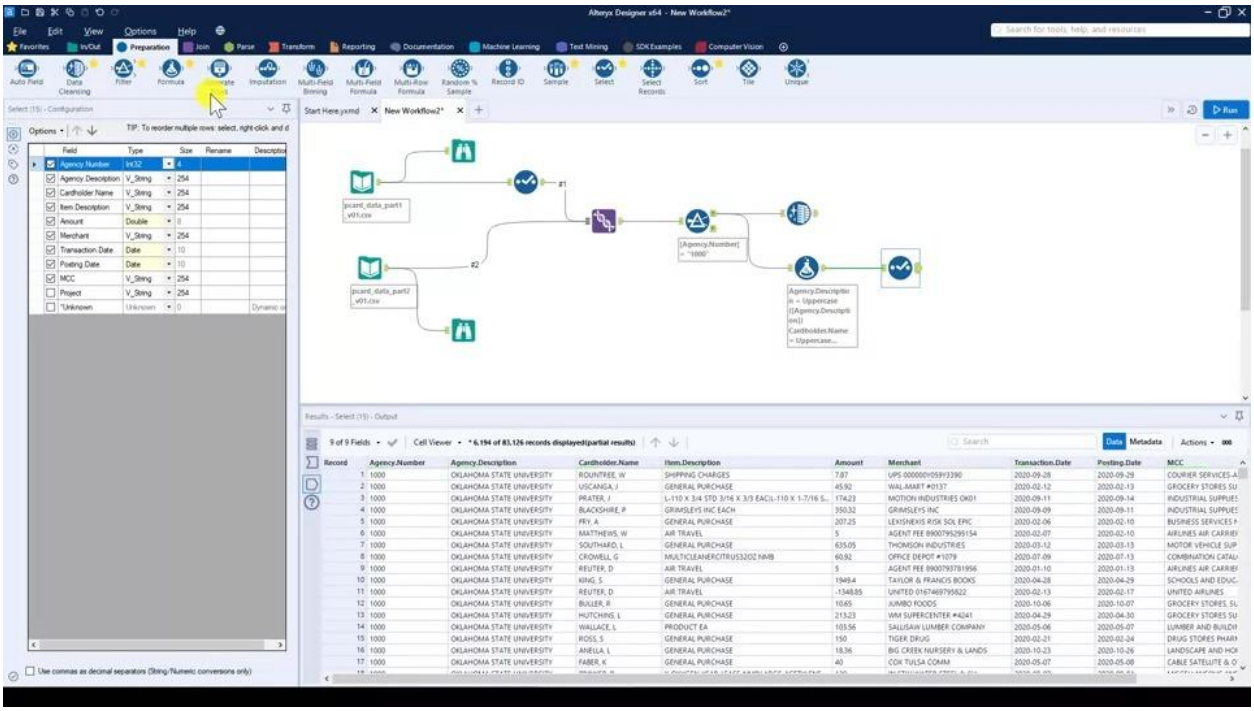
All right, so our next task here is to continue cleaning the data. And if we look at the combined data set, we see that we have two different agencies here. So we have the University of Oklahoma, and we have Oklahoma State University. Now, this case is about Oklahoma State University. So we don't want the University of Oklahoma. So what we need to do is filter that out, right? Or filter in only the Oklahoma State University. We also might notice that Oklahoma State University has agency number 1000. All right, so if we see that, we have a way to then get rid of the University of Oklahoma. So we need to filter this. If we go to our preparation tab, we see a very helpful looking tool in the palate, and that's the filter tool. Now, if we read through this description, that doesn't seem particularly helpful. But that is indeed what we want. Let's pull this over, and we're going to filter. And what we're going to filter then. If we go next to the configuration, we see this nice little basic filter that's already preselected for us, and it tells us to select a column. So let's do that. Let's make it easy on ourselves by selecting Agency Number, and we want that equal to 1000, right? So equal to 1000. Pretty simple. Let's go ahead and run this, and see if that works, [SOUND] okay? Now, we have two output anchors here. We have a true and a false. So the true is going to be if that statement is true. So this should be just Oklahoma State University, where agency number equals 1000. And sure enough it is. Just glancing through, we're only looking to subset of the data, but everything looks really good there. All right, if we click on false then, this should be when that statement is false. So this should be not Oklahoma State University. And the idea is we look at that output anchor, we can see under the false here, right? That we have University of Oklahoma. And so it looks like that worked so far so good. So our next task is to go in and to make sure we have the same case, upper case or lower case for all of the columns that we expect need to be

strings. So we have several columns. Some are definitely numbers such as amount, and some are definitely text. So one problem we might have and that we might see here is that we have different uppercase or lowercase mixed together in these files. So some here, this is sentenced case, this is all upper case. We need to have that be consistent. So all of our text columns need to definitely be the same uppercase or lowercase. Some tools will treat a name, for example, Roundtree here as two separate names if it has different case, and so we need to fix that. So there's a couple different ways to do this, I'll demonstrate too. If we go back up to preparation, the first tool in the palette that might strike our attention here, that might grab our attention is data cleansing. So this sure enough, it says, that does a lot of things, but it also modifies capitalization. So it's going to make the text uniform.



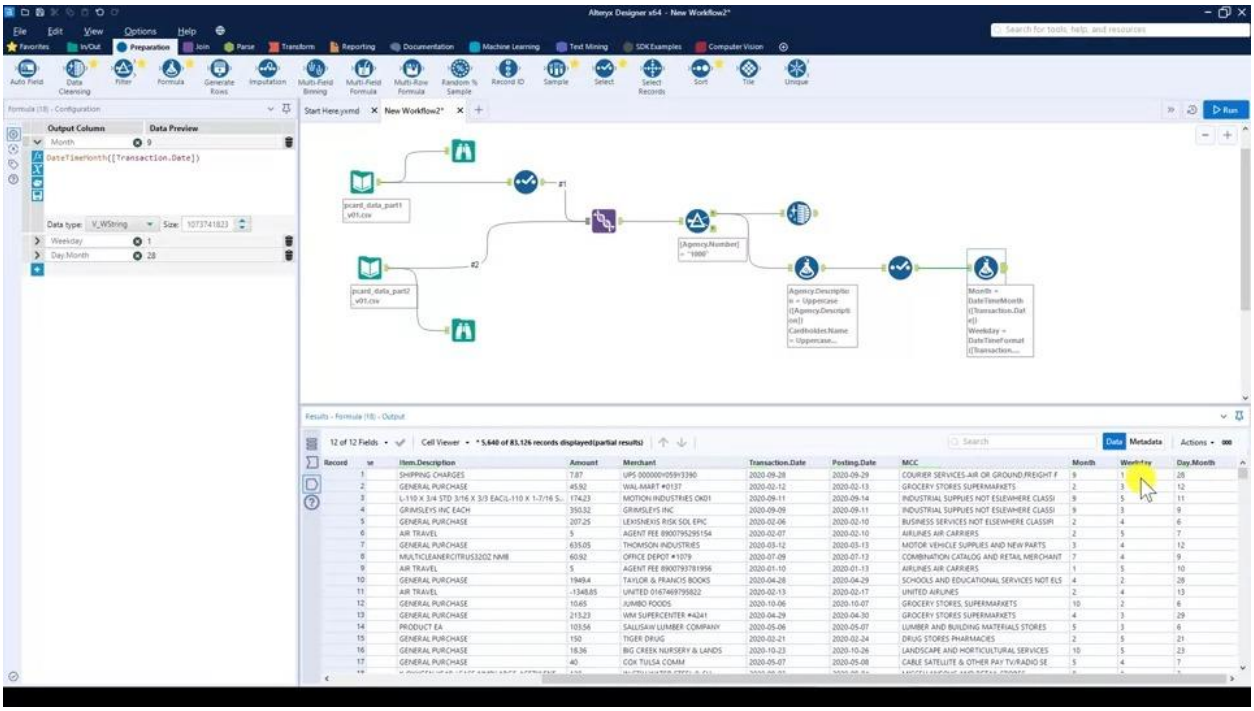
So let's bring this in, and we'll pull over again to true, right? Because we're keeping the true data, we're going to ignore the false. We only want Oklahoma State University. And if we look at this tool in the configuration area, we have a bunch of different things we can do. But one thing is we don't want to select everything, we just want to select the text one. So we have agency description, cardholder name, item description, merchant, and MCC. So let's pick those. Those are text files, the files that we want to end up being strings. And we're not going to do anything with nos. The only thing we're trying to do is change the case. So we only have that checked, and let's go ahead and make it upper case. So that is one way to do this. Another way is to use, a slightly more complicated way but a very powerful tool we could use to change this type of case would be the formula tool. You see here, as we click on it in the palette, this is performing a broad variety of calculations and operations. So let's drag this over, we're going to take from True again, of course. And as you can see over here in the configuration area, this is a little more complicated, right? We're going to have to first select the column, and then we have to select the expression of the operation that we're going to do for that. And if we click on this operation thing right here, this f(x) function formula, we can see that uppercase does show up in our suggested for this data type, that's what we're going to use. If that doesn't show up there for you, you can go down to string, and because we're operating on strings here, and go down to find uppercase. So here's our function, do an upper case on a string. And so the only thing that we need to do next then is to type in the column number. The column is surrounded by square brackets. So here, item description is what we want, and then we're done. So that's our first one. And we just have to repeat this process for each of the different columns that we're going to use. So there's agency description that we just did. Let's do cardholder name. I'm just going to

paste this in here, and then change the column because it's actually pretty easy to do that. And so the cardholder name. So if I just type of seeing here, it shows up cardholder name. And then you'll notice the nice preview that we get. So we can kind of double-check that everything's working as it should. The next one we'll go through and do is item description. And again, I'll highlight this, and just type an I, and item description comes in, and I'll click on item description. Couple more here, item description. Next, we have the merchant. So let's do that. And again, I'll just paste, and I'll type in an M here, [COUGH] and change this variable name. There's merchant. Again, our previews all look good, and we'll do MCC, which should be next. MCC, And that's it right there. So we've got 1, 2, 3, 4, 5. They all look good. We all see nice previews.



Let's go ahead and run this to make sure it works. [SOUND] All right, so if we click on the output from the data cleansing tool, right? We've got 10 of 10 fields. if we want, we scroll, and look at everything. All those texts seem in upper case. So that looks great. If we go over and we click on the output for our formula, again, same thing. We've got 83,126, 83,126. And again, if we click on our output anchor there, we can see things look really good. Everything is in uppercase just as we expected. So it seems that using both these tools worked, well and we're able to demonstrate both. So the next thing we want to do is continue cleaning the data. As you can see, if we look down at the data that we have now in the metadata, everything is still a string. Now, certainly, some of these are not strings, and they should be numbers, and some should be dates. So let's do three more tasks. First, let's go in and eliminate columns that we don't need, all right? That don't have useful information. Second, let's go in and change to numbers columns. It should be numbers, and change the dates columns that should be dates. We want to do this. The select tool is probably the best one for this. We've used this before to change names, and it'll work for this as well. Great tool, you can see over in the configuration area that we have a lot of great options. This is what we did before to rename the column. Let's get rid of this unknown thing, which isn't really a column but we don't need it, and let's get rid of project column, which is right here. [COUGH] Is not at all useful, and has all in a, okay? So we unselect project in unknown, and then let's go in and look at, then change our things in numbers. So we've got agency number, which is an integer, so there's no decimals. So we can pick one of these integer options. Any one of them is going to work. That just allows you to have bigger size numbers if you go to integer 64 versus 16. We'll just pick 32, it doesn't really matter for us. Next, let's go in, and we definitely need to change amount. We're going to use that quite a bit.

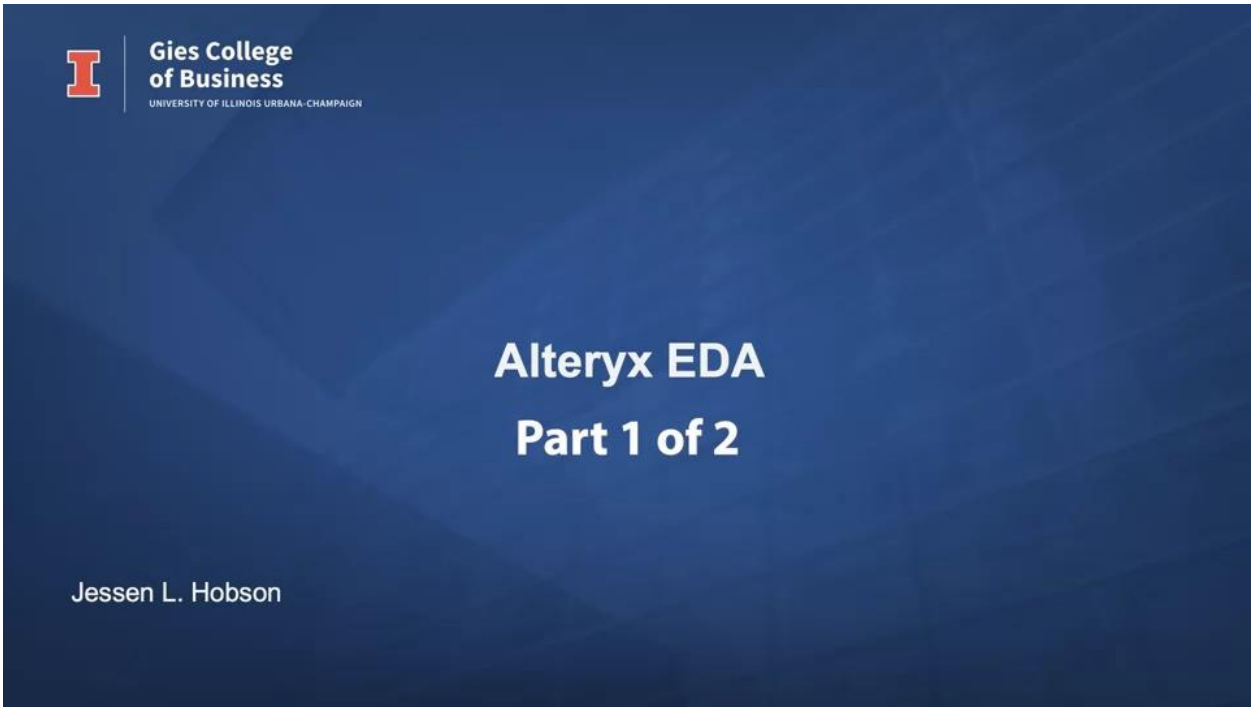
And so that is a decimal. And so there's different ways to represent a decimal. We can do flow, we can do double. Let's use double. That's kind of the standard, and give us the most options. So double is what we'll pick because amount is a decimal. And so there's our amounts. And then we just need to change dates. Again, there's multiple options too. We have two columns, their dates, and there's multiple options. Let's just take the simple one here, date. And date, and we'll run that. And we look then at the output anchor, and we look at the metadata, and we can see that things do look appropriate as they should, and if we look at the data, everything looks great so far. So done with that task. Okay, so for our last cleanup task, we're going to take the date column that we have, and to create three new columns, these are going to be really helpful for analysis later. And so let's use the transaction date column because that's the date that the transaction is actually happened. And what we want to do is create three columns. The first will be a column that list the month of the year. The second will be a column that lists the day of the week. And the third will be a column list the day of the month. And so we're going to do that again with the formula task. This is the one that's going to require us to write formulas for these.



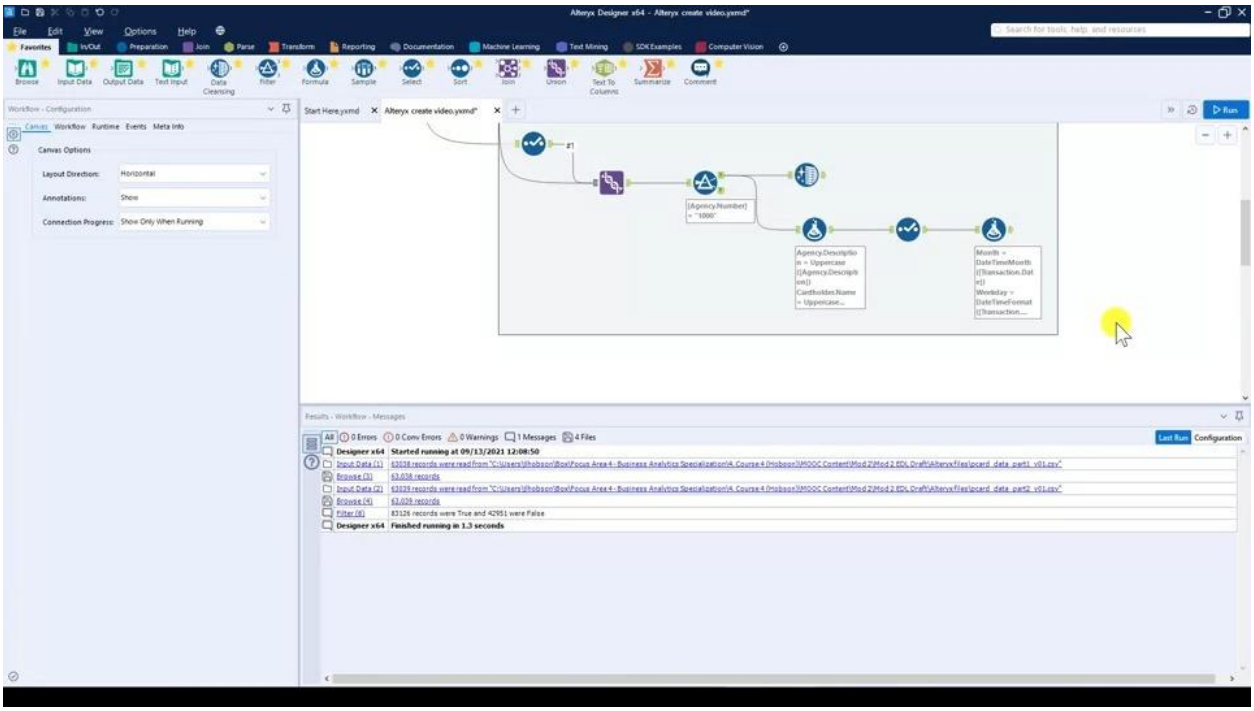
And so let's go over to the configuration area, and we have this nice setup, where it gives us the option to add a column. And so let's call this first one the month. This is the one where we want to list the month of the year, right? So this should be 1 to 12 for the first January to December. And so let's go ahead and try and see if we can find a function or formula for this, and there's nothing good in the suggested area. So let's go down to DateTime, which looks very useful and helpful. And let's see. If we go down to DateTimeMonth, we see that this return's a numerical value for the month of the datetime value, which is exactly what we want. The datetime value here are the column that we're working with, is going to be transaction date. So there we go, and the preview looks good. We're all set there. Let's go and create our next one. This second column that we're going to create, this is going to be again the list the day of the week. So let's call this weekday. And again, day of the week. This is going to be 1 through 7. And so let's look at that function and see if they have anything. And I won't waste your time, they don't have it. So this is something we're going to have to write on our own. And so what we need to do then is we need to use the generic version that they have. This is the DateTimeFormat. This is the generic function that allows us to do multiple different things. And so we'll click that, and again, we'll put in our column transaction date again, okay? And then this f is where we select what we want to happen here. And so here, I want to take percentage sign, w, and this is going to give us what we want. You can see so far, it looks good, and this one preview, it gives us, and we'll double-check it when we get to the end. And then we select our last column, and that's going to be, let's call this DayMonth. This again is a column that lists the day of the month. And so this is going to be 1 through 31. And so again, we go down to DateTime, and we say, DateTime. And this one looks good. So this returns the numerical value for the day of

the month. Exactly what we wanted, right? So we take DataTimeDay, and we again just add transaction date in there. And again, the preview looks great. And so let's go ahead and run everything, and we get no errors. And so that's great. And so if we look at the output anchor, let's look at our new columns and just kind of see if they make sense. So here's the transaction date is 9 28 2020, so the 28th of September, 2020. And so the month is 9, that's what we want. The weekday is 1. We wouldn't know that, but it looks like it's working correctly. And then the day of the month is 28, which is exactly right. And if we scroll down, we can see that this looks to have worked correctly. So excellent.

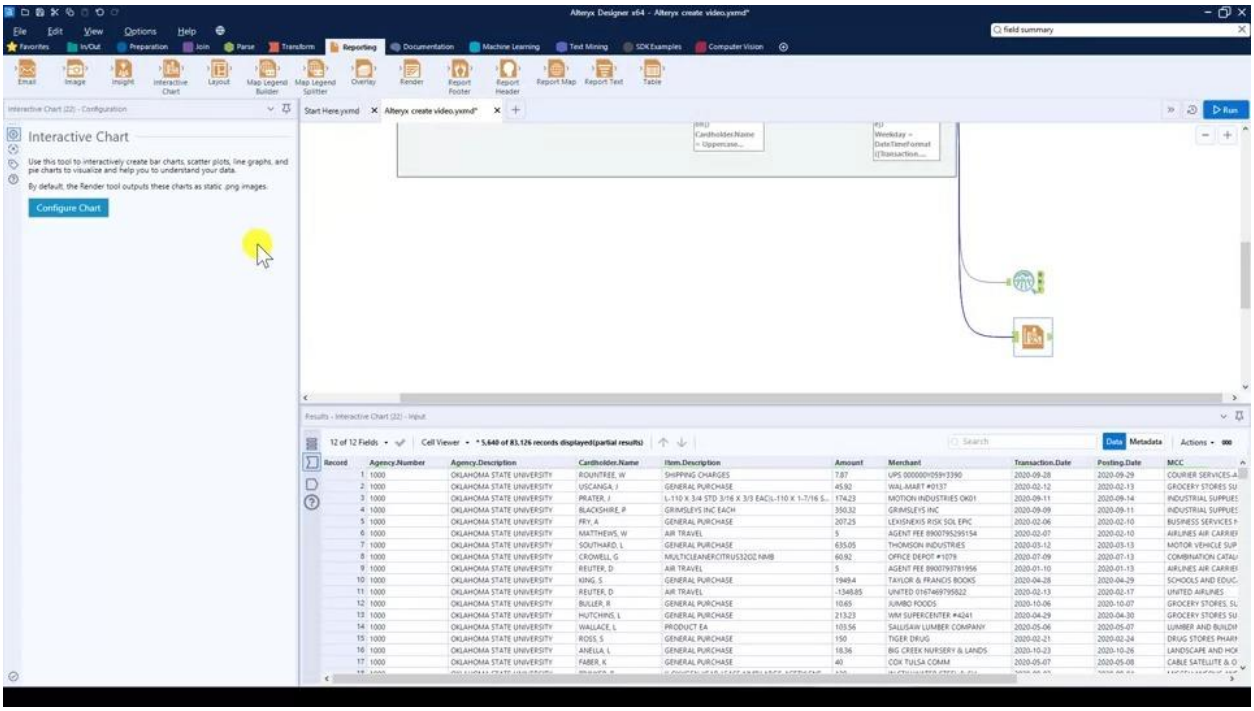
Lesson 2-8: Screen Shot Solutions in Alteryx - EDA Part 1



[Music]

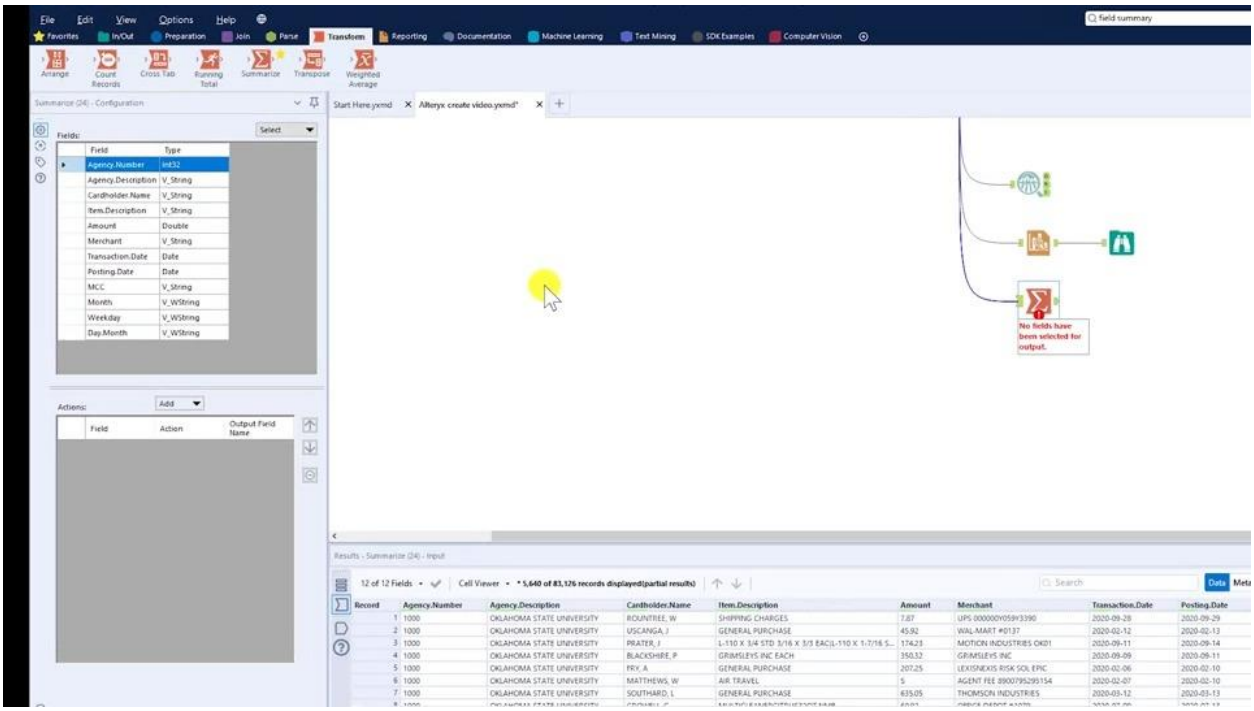


Okay, with our data newly cleaned the next step is to understand the data. In particular, to understand a little bit better about amount because that's one of the key things we're going to look at for our control test. So I think the thing we want to do is first probably organize this a little bit. And so I'm going to go ahead and highlight all of these and maybe drag them down just a little bit so that we can see things easier. All right, put that there. The next thing I'd like to do is demonstrate a container tool. This helps make things a little tidier as well as we start to have a bit of a larger workflow. So here I create a container I'm going to call this ETL Part A. And there we go. I can minimize this now and make it small so that it's easier to see what's going on in the workflow that we have. I can also do something that I will not do here but that if I click here I can disable the container, okay? And then it won't run at all. Now I won't do this here because we need all of our steps to run so that we can have the data. So that's one option there. If you have a piece of the workflow that you're not using, you can disable it. And so it won't run at all, saves a little bit of time. I'm going to rename this one as well over here in the configuration panel area as ETL Part B. All right so the next thing I'm going to do then is bring in a tool that will help me look at a summary of the amount column. And that's the field summary tool. It doesn't show up easily here.



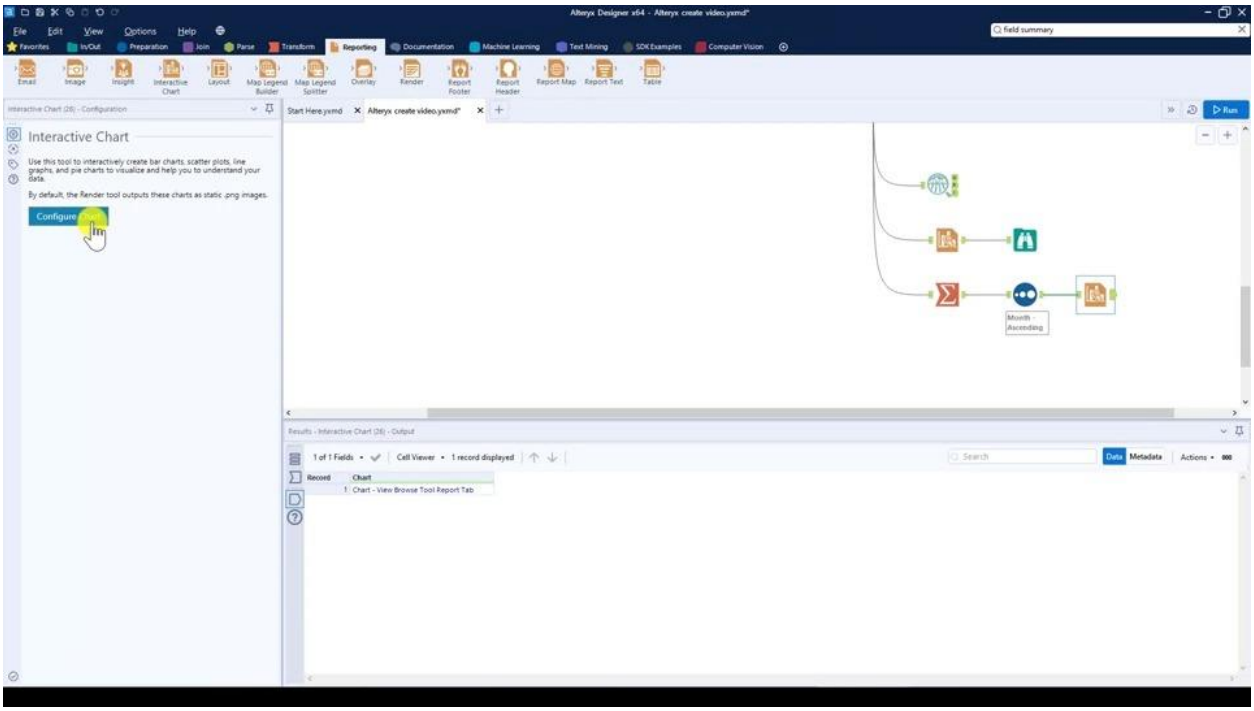
So I'm going to search for it. So I'll click on the search box and I'll type in field summary and this is the one I want and look this is nice. Another nice feature of all tricks. I can just drag this tool here. And it's going to work for me. I'm going to pull it down so that it's a bit lower here so I can work on things right there. Okay so I'm going to go ahead and pull that in and I'm going to look at it over in the configuration panel and again I'm interested in amount only. So I'm just going to click on a mountain and I don't want to sample and put down I want to look at all of it and I want to look at summary statistics for the amount column. And I'm just going to run and so we'll go through everything and run and once that's completed and you can see starting to take a little bit longer. I've got three options I've got the field of summary tool I've got option o which is output. I've got options r here which is the reports and I've got i. So with i and r you have to bring in a browser tool and we're not we don't need to do that. Let's just look at we'll click on o and we'll see some statistics. Some how statistics for the amount column you can see it named here. You can see it's numerous and it gives you these great statistics. There's several things that we want to point out here that they're interesting. First the minimum amount here if we look at the minimum amount it's negative. So this seems unusual you know as we contemplate what amount represents that is amounts the value of the purchases on the p card. Thus any negative amount must represent either a mistake or a refund of some sort. Certainly in addition to the control test we do in the future we're going to the internal auditor here. We definitely want to gain a better understanding of these negative amounts. Probably run some additional tests on them. The ardor would would inquire for management why the negative amounts exist and then test these amounts to verify what management tells them. Okay so the second thing we would notice here is that the maximum value is pretty high. Remember that we have a control

in place that the maximum transaction value for anyone transaction is supposed to be 5000 or less. So we already know just by looking at the summary statistics here that one of our controls has been violated. So clearly we there's a lot of work for the auditor to do here. So the next thing we're going to notice is the median and the mean. So here we have a mean of 87.65 and the mean of 258 and a half about. So this means the mean being higher than the median indicates that the data is right skewed. Meaning that there's several outliers on the high end of the distribution or in other words there's a few abnormally high values. So the order might do some additional testing on these high values by asking management about them. And gathering supporting evidence for them to examine and verify that they're valid transactions. So next let's continue to examine the distribution of the amount column by looking at box plot or a box and whiskers plot of amount. This plot is an easy way to graphically examine the media and inter quartile range of amount. It allows us to verify some of the things that we saw in the field summary tool. So let's go to report here and let's see if we find something that helps us out. And sure enough if we look right here we have interactive chart, let's go ahead and use this. We'll pull in and here I go pull that in from our clean data.

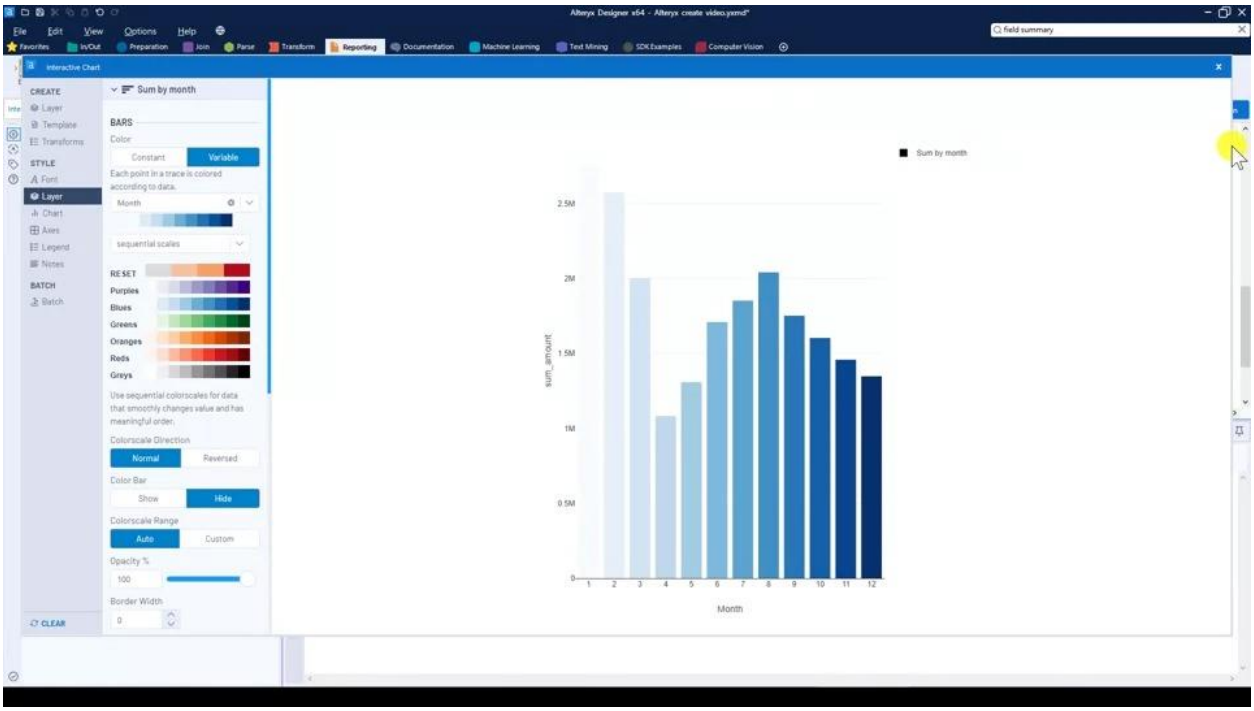


And so on the left here now in the configuration area you see this configured chart. And so this is nice, it's going to give us a bunch of different options to do and things to look at. So let's just start with, we'll add a layer, okay? And let's go ahead and we'll call this layer just called a box plot of amount because. That's what we want to do and then it gives us a type option and we want a box and whiskers plots will pick that. So there's that option and then x we're going to leave blank because we're just a box plot. So we're just looking at one variable and obviously want amount here. And so here we have a box and whiskers plot box plot and we can look at the amounts that we have here. So so the maximum is 23,000 as we saw before in the other tool that we just looked at. We can see the medians 88 which that sounds about right if I remember the meeting that we have this negative minimum. So things look great so far, although visually this is not very helpful, is it? So what we want to do is let's go ahead and scroll down to x. Okay let's go over here two axes and change the display range there it is 4y? So let's customize this and let's just set it to like negative 1000 and 1000 and see what that does for us. See if that makes things look a little bit better and indeed it does. Again, if we hover here, we can see the same amounts from before. But this is this is pretty helpful. We can see the rights que nous as we look at the bigger area here above the median. And so that's actually quite helpful. So let's go ahead and get out of this and there's no good way to view this unless we keep clicking on configure chart or we can output it. So let's use our brows tool right here. And so if we run this then we'll create that chart the way we just configured it will take a minute of course. But we can go and click here and we have our chart over here in the configuration area and we can pull it over and there we go. There's our chart. Next let's visualize the amount of transactions that happened each month. That is let's sum up the transactions for each month and

visualize that. To do this we'll need four separate tools. So first we'll need the summarized tool because we can we want to summarize amount in two months, we want one entry, one, total sum of all the transactions by month. We can do this with a summarized tool. Next let's sort months from 1 to 12. So we can visualize the data by month. After that, we'll use the interactive chart tool again and then we'll use the browse tool again so we can visualize things more easily. Let's go ahead and move this down. And let's go grab the summarized tool. Let's see, so that's going to be under transform, here's summarize, okay? So that's what we need to pull this over. Let's go ahead and bring the data in our clean data in.

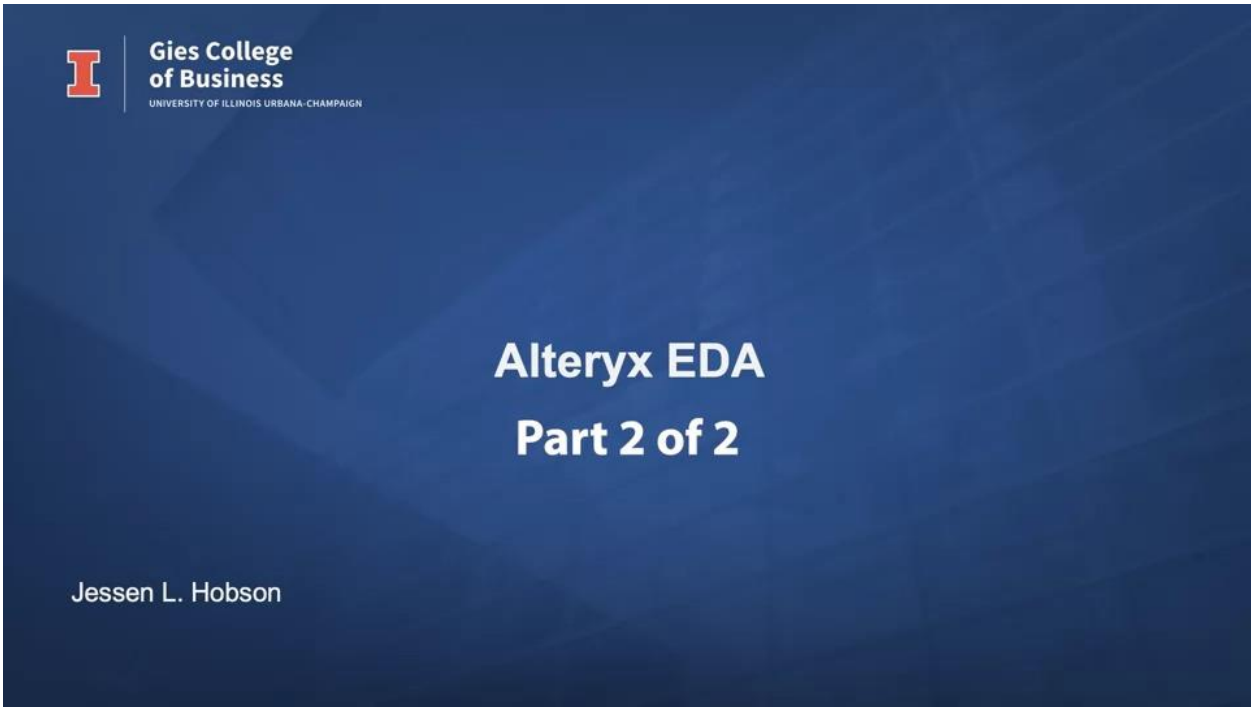


All right, let's go over then and total amounts for each month. So we want to first group by month. So let's click on month right here and that's why we created that variable. Let's add a group by here. So they were grouping by month. That's what we want. And then we want to sum up amount within each month. So let's click on amount and we'll add a some and we'll rename this to some amount. I just make it lower case. You could call this whatever you wanted. Okay, so that's done there. Our next step then, is to sort and we want to sort by month from 1-12. So that's in our preparation area. So we go to the preparation area, we pull down sort and we're just going to sort by month ascending yeah, so 1-0. The next thing that we're going to do is we're going to use our interactive chart tools. So let's go back and find that under reporting. And here we grab it from interactive chart, we pull that over. All right, so let's go then over to our configure chart. Again, we want just to look at I think a bar chart would work well here. So let's just call this sum by month. And let's try a bar chart. We think that look good. So the X axis is going to be month, right? Because we just want to look at one bar, the height of each bar for each month and then we want to use sum amounts. And so that will make sense. And before we can change that we'll have to run this. So let's go ahead and run that so we can get the data all the way down here so that we can get it sorted and ready to go for our tool. It's nice that we click on the tool here, we have no errors at all. So that's great. And so things appear to be sorted. And if we kind of look at the data we can make sure. And let me pull this up just a bit and we can see that sure enough, we have each of our 12 months and we have the sum of the transactions for each month. Again, a reason to use the visualization tool is because this gives us the data but it's not super easy to see and understand.

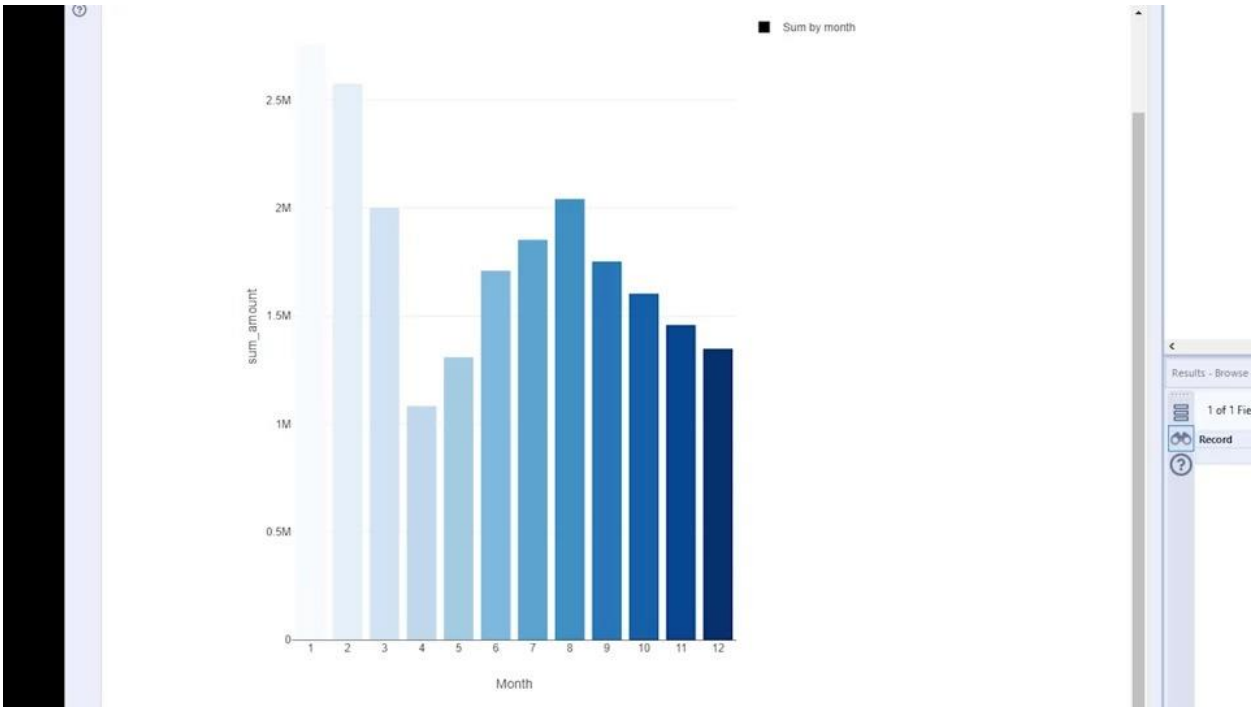


So let's go back to our configure chart and so that looks pretty good, right? And this is much more discernible now, we can see quite a few things and draw some conclusions from there and maybe give some steps for the auditor to take in the future. One thing that might be useful is if we go in here to layer and let's kind of put a different color for each month. So if we do that, if we go into variable, we can do this. We can select a month and, select some interesting color scheme that we want. You don't know what you're going to pick. You can pick whatever you want, maybe blue, whatever your favorite color is and do that. Okay, so that's done.

Lesson 2-9: Screen Shot Solutions in Alteryx - EDA Part 2

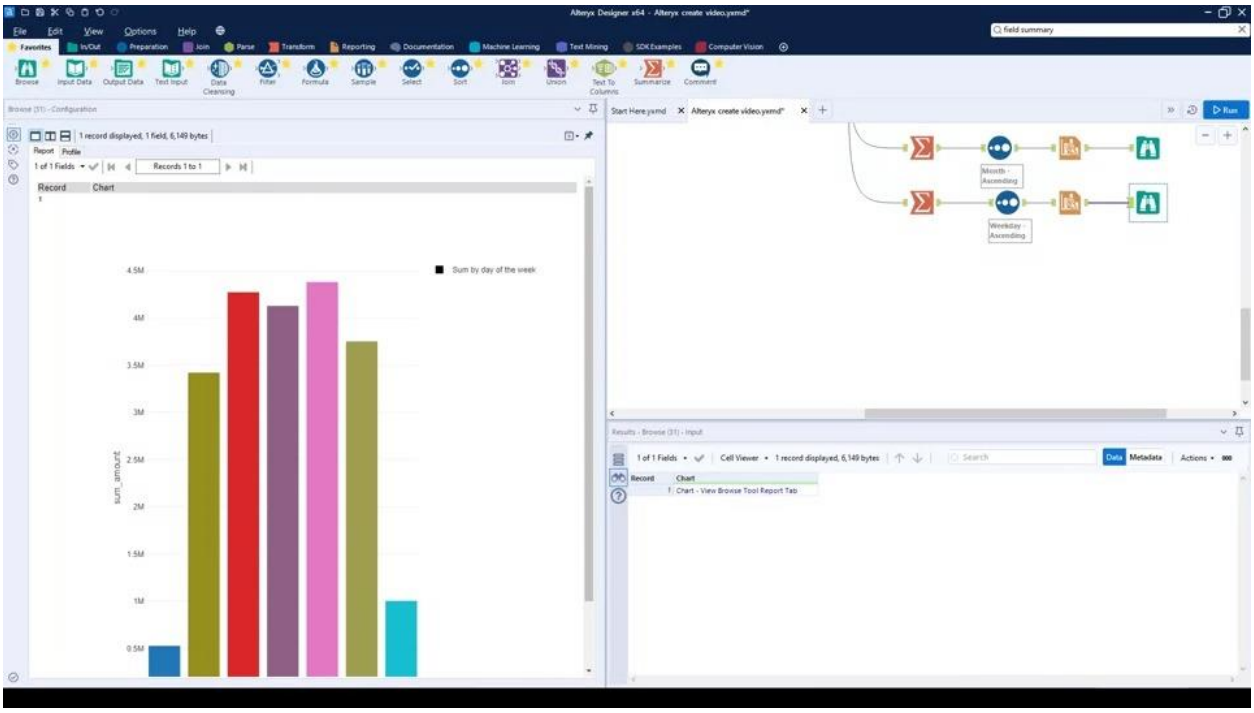


[Music]



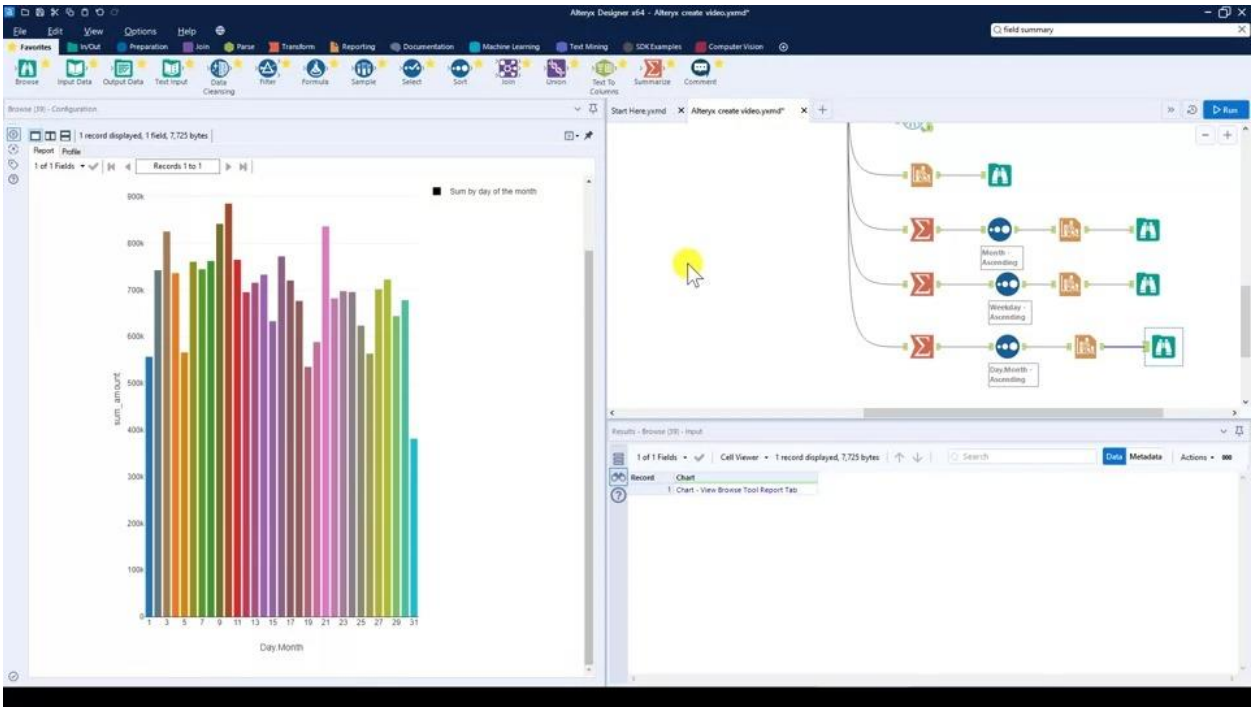
And let's next start our brows tool and we pull over brows and let's run this again so that we can look at it. All right, so once that is run, just drag this over and we can take a look. So okay, what do we learn from the graph first, what probably stands out is that most of the transactions really happen in January and February. Those are super tall here, right here on the left of the graph. What, this should definitely be of interest at the order. The order would want to know why this is the case. It's not immediately clear to me why January and February would be the most active months for p card purchases at the university. It might be that more transactions happen at this time or that they're of the transactions that do happen or of a higher amount to higher value or some combination of both of those. Certainly January is the start of the spring or winter semester at most universities. But it's often the case that the fall semester which starts in September is the busier, more active time of the year for university and for purchases. So perhaps the university is making a lot of capital expenditures, building materials and whatnot at the beginning of the year to fit with state budgets. It might be that the beginning of the calendar year is a time when the university makes a lot of purchases of hotels for air travel for conferences, for example. Well anyway, more consistent with our expectations is the high amount of transactions in august. All right, that makes sense as being close to the beginning of the fall semester. All in all this graph does provide several new questions to the auditor that they need to ask the management. And indicates several additional tests that the auditor needs to take and do. Specifically the auditor could search the descriptions of the item description column or the merchant column of the MCC column there in January and February to kind of find out what's going on. They might search for terms about construction or air travel to see if that's what's driving the high number of purchases in January and February. For

our next visualization let's use the week day variable that we created earlier to visualize the summer purchases made based on each day of the week. Once again we'll need the summarized tool and then the sort tool and then the interactive chart tool and then the browse tool.



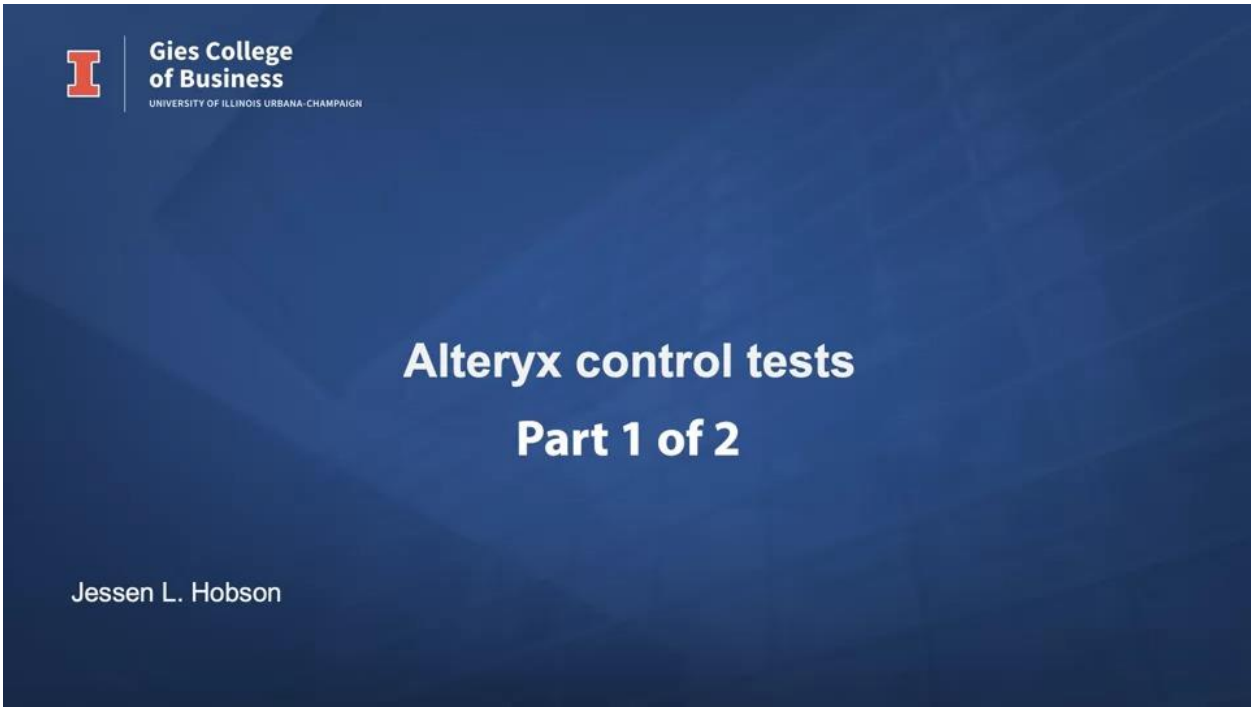
So let's go ahead and put this back. I had to bring each one of those out individually. So, I've got these now on my favorite, so I'm going to bring out the summarized tool. Let's go ahead and bring the sort tool in here now than an interactive chart tool and our brows tool once again. Okay, pull on the data. So now again our goal is to we're going to group by a week because we want to look at the day of the week rather we're in a group by weekday because we want to see one observation per a weekday. So, let's go ahead and pick that one. So we're going to pick week day, which is one of the new variables that we created, select that here. And then we'll add a group by group by that and then we're going to again sum up the amount and so we'll add that. So here's are some and I'll change the name. Just make this lower case. Just for my preference. You can do whatever you want and sort by a weekday ascending again as we did before. And then we'll do the first part of our configuring this chart. We're going to once again at bar chart here. So it is the sum by the of the week. All right and we'll pick the bar chart. We did this before. Well x is going to be the weekday and then why is going to be some amount and so let's go ahead. Run this and see what we get, make sure that there's no errors at all. We do notice from before the interactive chart here. There's a warning that this is back with the bar chart that the dataset is greater than 110,000 rows that can cause a slow chart rendering. We're not going to modify that. We don't care. We're not going to sample. Hopefully your computers are quick enough to work with that. So no air is that warning we're not going to worry about. Let's go ahead and take a look at our

chart here, pull this over and this looks pretty good and I think we can learn some things from this. Let's change it up just a bit. If we can, let's go over here and go to our chart let's just change the colors. Let's color again as we did before by month. So if we go into layer and then we click on variable and we select the week day we can change this around. Look at all these different options we could pick,, so just play around with these tools against go ahead and run this so we can see how that looks in our brows tool. So what do we learn from this plot first? It's probably obvious just from looking at it. But it's important to note that weekday equal to 0 is Sunday and weekday equal to 6 is Saturday. Thus, from the graph we can see that the majority of transactions happen as we would expect during the normal five weekdays Monday through Friday. But probably the biggest finding here is that there's a half a million transactions on Sunday and on Saturday we have another whole bunch of transactions over a million transactions on Saturday. So all in all,, we'd expect a priority that there would probably be some transactions on the weekends. For example, academics frequently travel on the weekends and hold conferences on the weekends. However, 1.5 million spent on weekends over the course of the year seems potentially like a lot. So the auditor would definitely want to do some additional testing on transactions that happen on weekends and we certainly inquire about these transactions to management. Okay, last, let's visualize the sum of transactions by the day of the month. So maybe we have different transactions, a different behavior at the beginning of the month versus the end of the month versus the middle of the month. So let's visualize that and see if we find anything interesting.

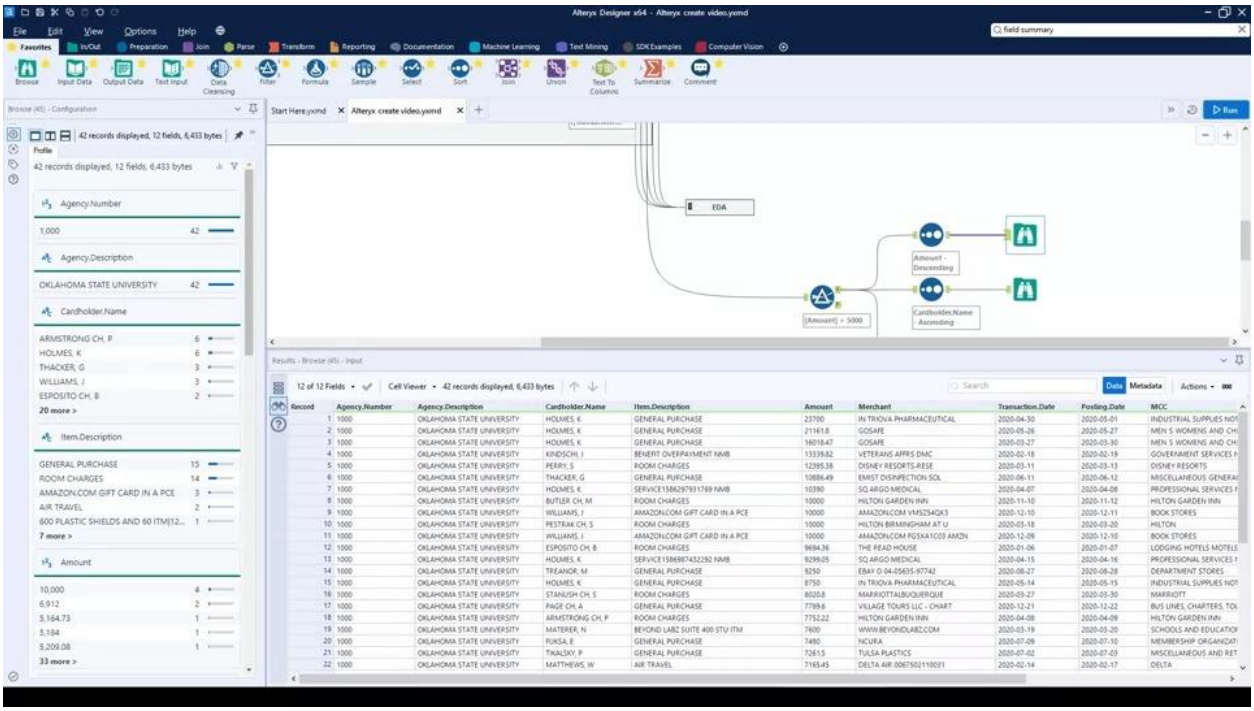


Again, we're just going to use all the same tools here. Again, I'll just kind of throw them all together at the beginning. Once more. Like this, pulling the data as we've done in the past. Okay, now we're going to summarize by that knew that last new variable that we created, which was date a month and so well grouped by date a month. Again, this time and we'll once again take amount and sum by amount. Once again, we'll change this to lower case just because that's my preference. All right, so that's done sort by day, month, let's configure our tool again to use a very similar chart will add our layer and this will be called some by day of the month. And again, bar charts and we'll click here a day of the month and so some amount. And so that will be our chart will run that and then let's go change the colors again because we like to do that. No errors. Which is great with that one morning, which are not worried about. And we go in here and let's go down to the style and we'll change the layer and we'll add that variable stayed a month again. And let's go ahead and change this to something interesting. Yeah, how about that? All right, run this one more time so we can look at it again in our brows tool. So here's our brows tool, here's the chart. So now do we learn anything from this plot? Not really the lowest transactions appear to be at the beginning and the end of each month with peak purchase days scattered throughout the course of the month. None of this seems overly surprising. Perhaps a more thorough investigation would examine how these vary by month. In particular, the author might be interested in January and February and august since those were there really active months that we saw in our last graph. All right, so there's some good visualization of our data and we're ready to go in and do some control testing of it.

Lesson 2-10: Screen Shot Solutions in Alteryx - Control Tests Part 1

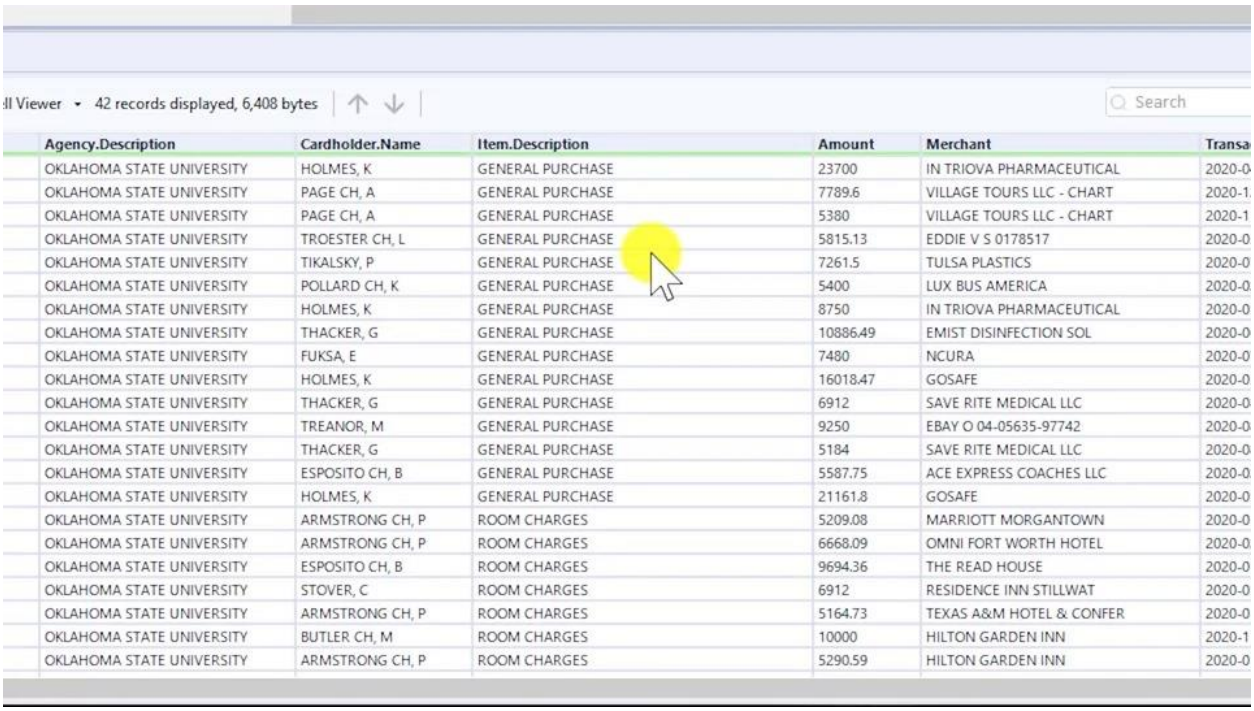


[Music]



Now in this video I'll perform simple tests of controls for each of our three stated controls, over purchase amounts for the Oklahoma State University p card transactions. So there's three specific controls, the first one is that total purchases for the year are less than 50,000. So that is an employee is not allowed to make purchases totaling more than $50,000 over the course of one year. The second control is that employees are not allowed to make purchases totaling more than $10,000 a month.. So $10,000 a month, and the third is that individual transactions are not allowed to exceed $5,000 each. So before we do that let's tidy up a little bit, and so we'll drag this back over here, pull this down just a second. I'm going to make a container here, that I'm going to use and I'm actually going to make it inactive. Because I don't have to run this every time and that might speed things up just a little bit. So here's a new container, now I'm just going to call this EDA over here in the configuration window, like that. And I will inactivate this so I'll disable this container by clicking here, okay. So let's start with the third control that individual transactions must be below $5,000. To do this, let's first filter all transactions so that we're seeing only the ones that are above $5,000, and let's see what that gets us. Let's sort these transactions in three separate ways, first we'll sort on amount based on highest to lowest or highest to lowest based on amount. Second we'll sort by cardholder name, and third by item description. So what we're going to use then the tool will use next is filter, so because we want again to filter see this is going to query records based on expression split that into two streams. So we use this once before, and so we're going to pull this in here, and let's put it over here and we'll drag the data to it. And again, what we want to do is filter here based on amount, right? And we want to have a mount be above $5,000, so we'll put greater than 5,000. Pretty easy, and that's really all we need to do so we've done that, and then the next thing we're

going to do is we mentioned sorting in three different ways. So we're going to pull this based on true and we'll just do four, three separate sorts, There's a second, And here's a third. All right, so this first one we'll go based on amount, And just sort that descending, so that we have the highest first. The next we'll do cardholder name ascending, so here's cardholder name ascending. And the third look at this, let's go ahead and do item description, Ascending. All right so we'll do each of those, pull this down, let's pull over a browse for each of these so we can see it, and then we'll run it and start looking at it. Lets first after we run things, make sure that we have no errors, It looks like things are in order with no errors. So what do we learn from this test of control? Well first we see that there are 42 violations over the course of the year. Now this is a small percentage of the total transactions that happened, so that's comforting to the auditor. But the auditor will need to discuss each of these exceptions with management and find test management's explanations, so whatever management says, the auditor will need to test that. The auditor has at least two goals and testing these transactions, the first is to make sure that the controls are not violated. So here we clearly see there's 42 violations, the auditor will need to understand why these violations have occurred. And if there is some explanation for the violations, and then whether changes need to be made to the controls to prevent violations in the future. Of course it might be the case that these violations are in fact warranted and appropriate. In that case, the auditor will need to gain a better understanding of what the controls should be, and when they can appropriately be violated, and changed. The second goal in evaluating these controls is to understand the possibility of the financial statements are being misstated. That is the controls are in place to prevent inappropriate transactions. And the auditor needs to gain comfort that these transactions are not misstatements of the financial statements are issued by Oklahoma State University. Let's look next at the browse tool for transactions sorted by amount, so we'll click here, pull this up a bit. What do we learn from this view of the data?



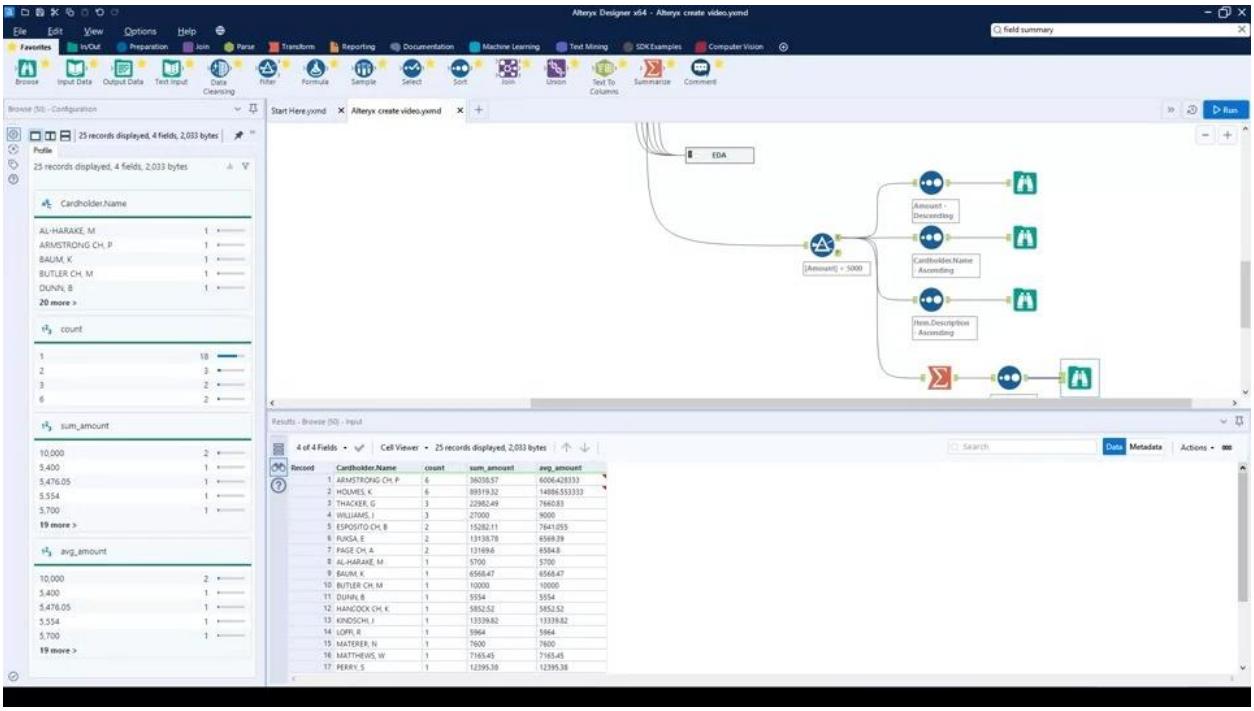
ll Viewer ▾ 42 records displayed, 6,408 bytes

Search

Agency.Description	Cardholder.Name	Item.Description	Amount	Merchant	Transa
OKLAHOMA STATE UNIVERSITY	HOLMES, K	GENERAL PURCHASE	23700	IN TRIOVA PHARMACEUTICAL	2020-0
OKLAHOMA STATE UNIVERSITY	PAGE CH, A	GENERAL PURCHASE	7789.6	VILLAGE TOURS LLC - CHART	2020-1
OKLAHOMA STATE UNIVERSITY	PAGE CH, A	GENERAL PURCHASE	5380	VILLAGE TOURS LLC - CHART	2020-1
OKLAHOMA STATE UNIVERSITY	TROESTER CH, L	GENERAL PURCHASE	5815.13	EDDIE V S 0178517	2020-0
OKLAHOMA STATE UNIVERSITY	TIKALSKY, P	GENERAL PURCHASE	7261.5	TULSA PLASTICS	2020-0
OKLAHOMA STATE UNIVERSITY	POLLARD CH, K	GENERAL PURCHASE	5400	LUX BUS AMERICA	2020-0
OKLAHOMA STATE UNIVERSITY	HOLMES, K	GENERAL PURCHASE	8750	IN TRIOVA PHARMACEUTICAL	2020-0
OKLAHOMA STATE UNIVERSITY	THACKER, G	GENERAL PURCHASE	10886.49	EMIST DISINFECTION SOL	2020-0
OKLAHOMA STATE UNIVERSITY	FUKSA, E	GENERAL PURCHASE	7480	NCURA	2020-0
OKLAHOMA STATE UNIVERSITY	HOLMES, K	GENERAL PURCHASE	16018.47	GOSAFE	2020-0
OKLAHOMA STATE UNIVERSITY	THACKER, G	GENERAL PURCHASE	6912	SAVE RITE MEDICAL LLC	2020-0
OKLAHOMA STATE UNIVERSITY	TREANOR, M	GENERAL PURCHASE	9250	EBAY O 04-05635-97742	2020-0
OKLAHOMA STATE UNIVERSITY	THACKER, G	GENERAL PURCHASE	5184	SAVE RITE MEDICAL LLC	2020-0
OKLAHOMA STATE UNIVERSITY	ESPOSITO CH, B	GENERAL PURCHASE	5587.75	ACE EXPRESS COACHES LLC	2020-0
OKLAHOMA STATE UNIVERSITY	HOLMES, K	GENERAL PURCHASE	21161.8	GOSAFE	2020-0
OKLAHOMA STATE UNIVERSITY	ARMSTRONG CH, P	ROOM CHARGES	5209.08	MARRIOTT MORGANTOWN	2020-0
OKLAHOMA STATE UNIVERSITY	ARMSTRONG CH, P	ROOM CHARGES	6668.09	OMNI FORT WORTH HOTEL	2020-0
OKLAHOMA STATE UNIVERSITY	ESPOSITO CH, B	ROOM CHARGES	9694.36	THE READ HOUSE	2020-0
OKLAHOMA STATE UNIVERSITY	STOVER, C	ROOM CHARGES	6912	RESIDENCE INN STILLWAT	2020-0
OKLAHOMA STATE UNIVERSITY	ARMSTRONG CH, P	ROOM CHARGES	5164.73	TEXAS A&M HOTEL & CONFER	2020-0
OKLAHOMA STATE UNIVERSITY	BUTLER CH, M	ROOM CHARGES	10000	HILTON GARDEN INN	2020-1
OKLAHOMA STATE UNIVERSITY	ARMSTRONG CH, P	ROOM CHARGES	5290.59	HILTON GARDEN INN	2020-0

First, we can see that some of the violations of the 5,000 spending limits are large and significant violations. For example we see that K Holmes, here has made purchases of $23,700, 21,162, and 16,018. So these are all for something called general purpose, if we look over the item description. Thus the auditor will want to make sure that these transactions are appropriate, and that they should not have been made in some other way. Next let's look at the view of these transactions sorted by cardholders, so we'll click here. This analysis is interesting because, while we see a few individuals who have only one of these transactions, we also see several individuals that have more than one transaction. For example here we have Armstrong Ch, they purchased charges at various hotels at above $5,000 about six times. And then we have we have K Holmes down here, who purchased several things above the limit. Williams, also purchased items, Down here at the bottom, Amazon gift cards so three separate times in early December. Again all of these need to be investigated by the auditor and this particular sort of the data illustrates that some of these individuals may have roles that do allow them to make purchases above the transaction amount. But this is new information for the auditor that they need to come and understand. Furthermore the amazon.com purchases seem a little suspect and the auditor should be skeptical about those, and understand what's happening there. Perhaps they're approved for christmas or holiday gifts for the staff for example. On the other hand, they might be personal purchases that are inappropriate and in violation of controls. Finally, let's look at the view that sorts the data by item description. This view of the data provides additional insights into what types of transactions are likely to violate the individual transaction limit of $5,000. The first one here is pretty interesting yeah, this is 600 plastic shields and it's almost certainly related to COVID-19, and it's probably not a problem. We then see a large

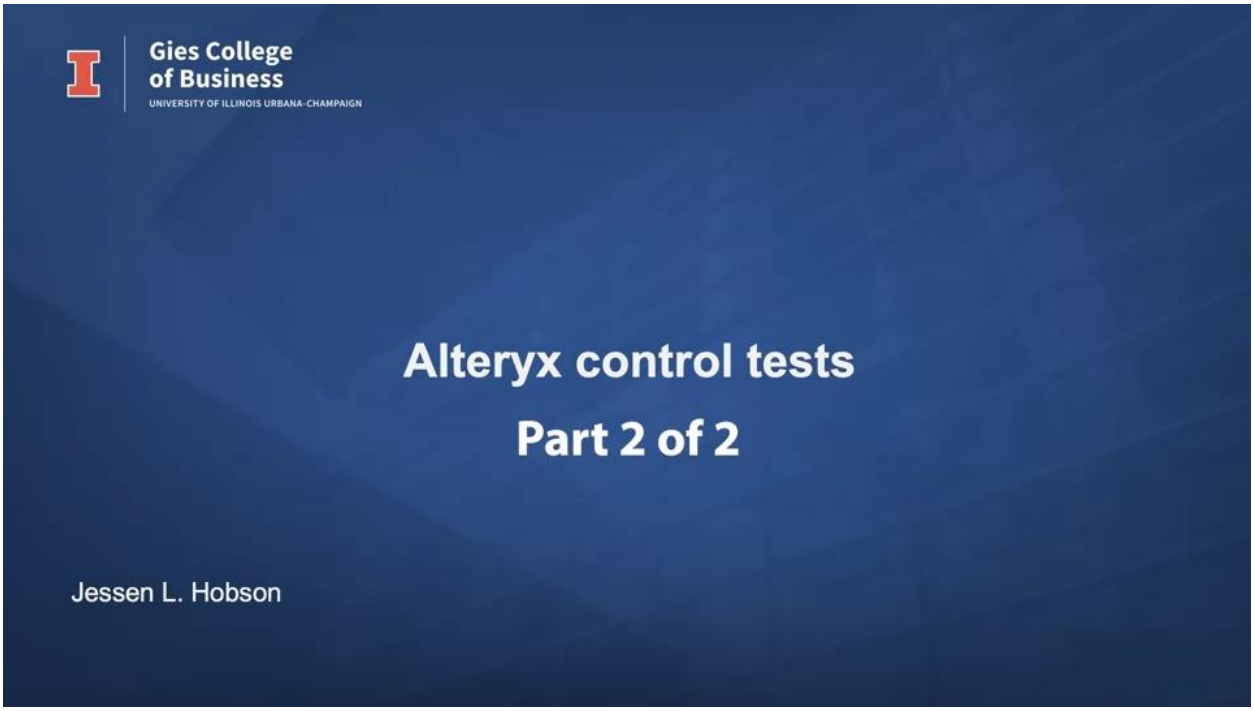
traunch of air travel and room charges, and these are probably appropriate. They probably relate to academic travel and to travel to conferences, and it does seem like the same individuals generally make these purchases. So the auditor just needs to understand when these charges happen, and if they're allowed to exceed the $5,000 limit. Now the most concerning category is this one here with general purchases. The auditor really needs to know when and why these purchases are allowed and what they are and why they're allowed to exceed the $5,000 limit. They seem to be for various items such as bus travel, uniforms, lab equipment, etc. So in general the auditor should be extra skeptical about anything that has a generic label such as general purchases.



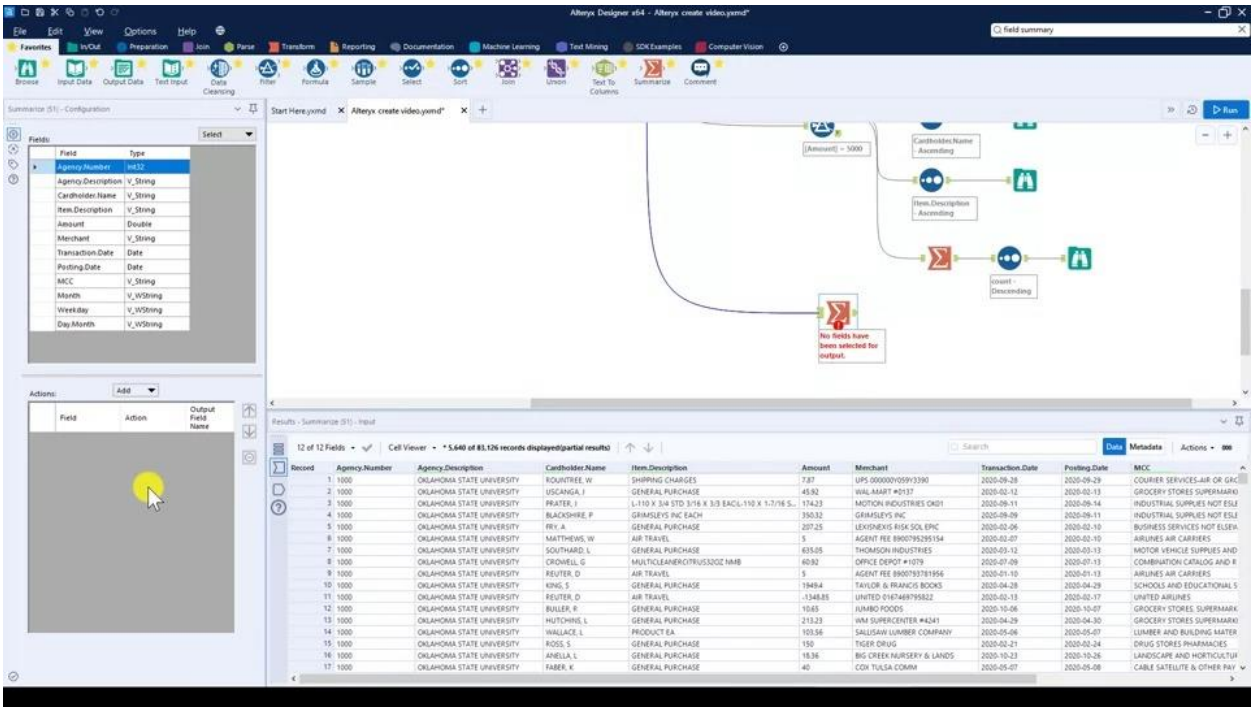
All right, let's do one final analysis here of the $5,000 individual transaction control. Let's group all of the violating individuals together and do three things, let's first count the number of violations for that individual. Second, let's sum the total dollar amount of violations for that one individual. And third let's calculate the average transaction amount for each of the individuals violations, then let's sort those findings by descending amount. All right, so let's go up and let's drag down the summarized tool okay, we can put that right here. The data we're taking is for those transactions that are greater than 5,000, all right? And again, we're going to do several things here first, let's group by cardholder name. So this again is an analysis that looks at individual violations, but does so by the particular person. So we group by cardholder name, then let's do some analysis on what we find here. First let's count up those transactions, so here we're going to count, use them out and count I like to make these lower case so let's do that. So let's call this count next, we'll do a similar thing and sum so sum the total by person, okay? And finally we'll do one more, and we'll take the mean the average, numeric there we go. Okay so it's under numeric and we'll just call this average amount, let's then go in and we're going to sort based on count. There's multiple ways we could sort but let's use count, we'll pull that over and sort by count. And then let's take a look at that so let's run this, and then look. [SOUND] This is another useful view of our data first, we learned that a couple of employees have a six total violations of the $5,000 transaction limit. But that most of the violations are just single transactions from one person. So looking at the sum of the amounts, we see that some of these are quite high, such as $89,000 for K Holmes. Further, we see that each of Holmes' violations are an average of $15,000. So again this individual has six total transactions that violate the $5,000 limit. The sum of all of those violations is $89,000,

which is high, and the average amount of each of those transactions is almost $15,000. Three times the amount of the $5,000 limit so, clearly this individual needs to be investigated more thoroughly by the auditor. The sum amount here this column again then clearly indicates that our other controls, such as the limit for spending for one year are being violated, and so we'll have to look at those next. So the control that we'll investigate next is the yearly spending limit control. That is total purchases for the year for one individual must be below $50,000. So our test of this control will be just to group and summarize all transactions by cardholder name, and then keep only some transactions are above $50,000. Let's also at the same time count the transactions for each individual. So this is definitely one area where it's important to make sure that the cardholder names are in the same case, either all uppercase or all lowercase. So that all tricks or another tool we might use later, does not classify a person with the same name as two separate cases as two separate people, so it's good that we did that analysis before.

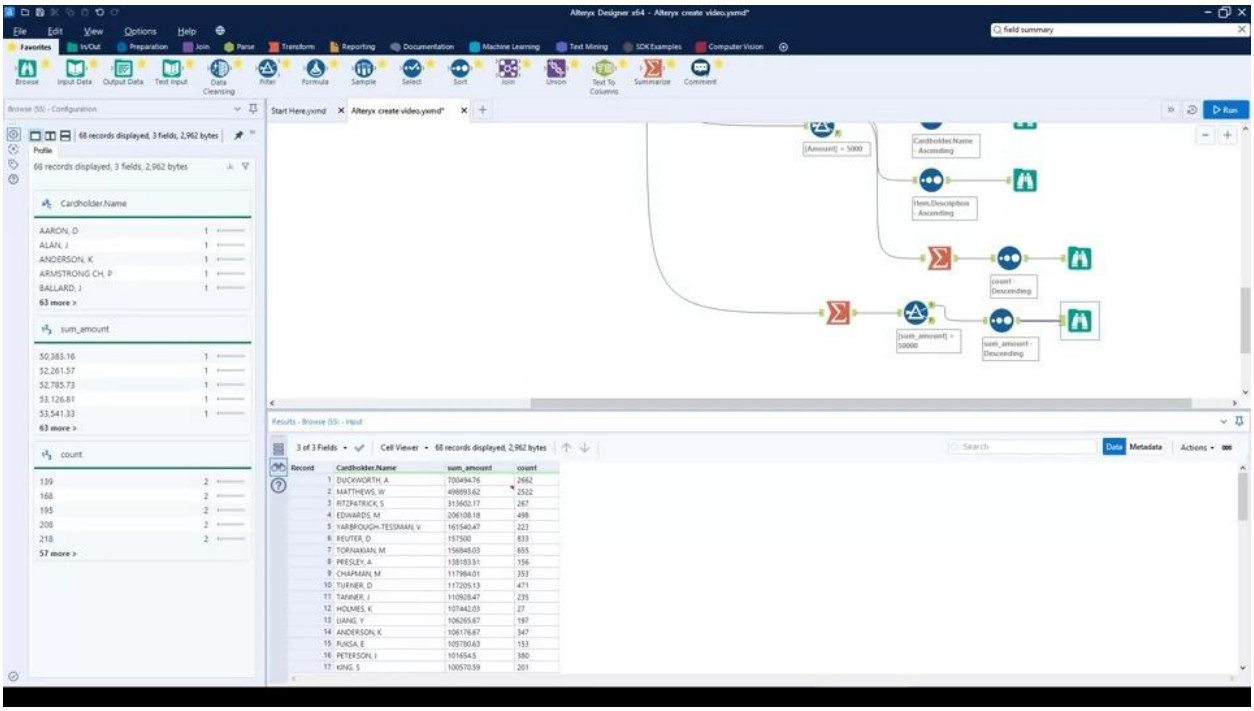
Lesson 2-11: Screen Shot Solutions in Alteryx - Control Tests Part 2



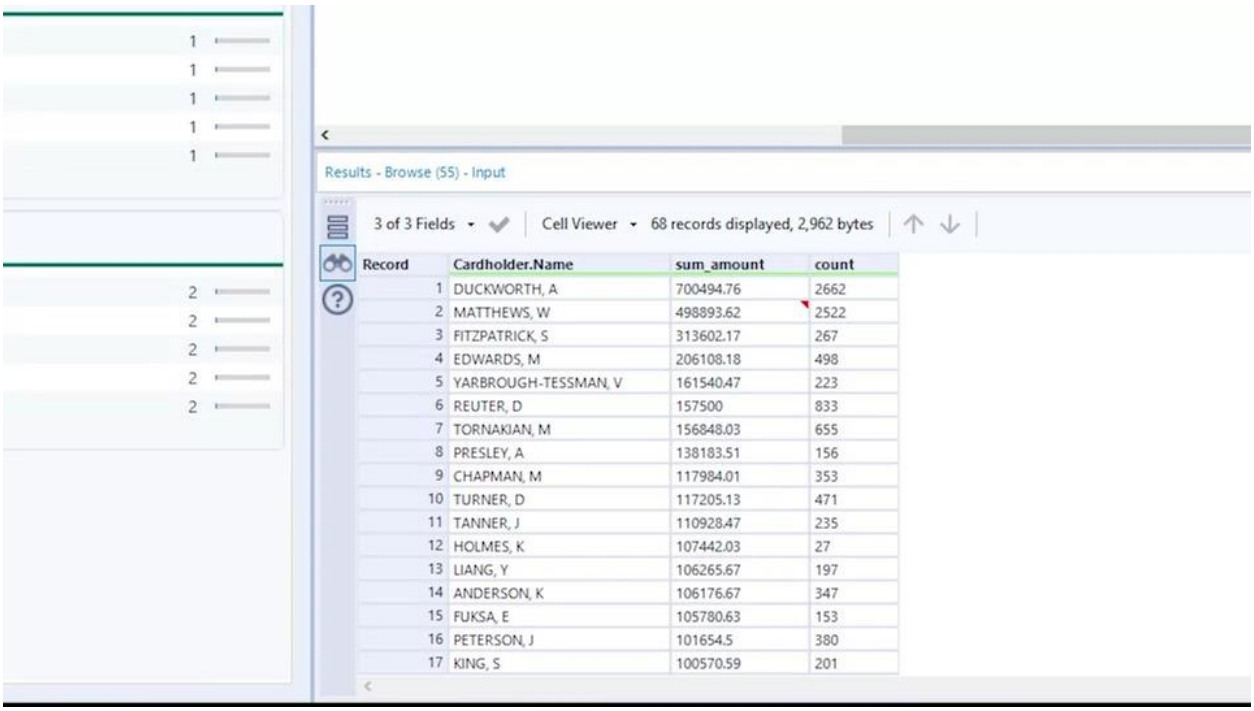
[Music]



Let's go ahead and do this. First, what we need to do is summarize, and this is because what we're looking at here is a summary by the individual for the year. We're going to summarize, so let's pull in that data all the way from up here. There we are, and let's go over here.



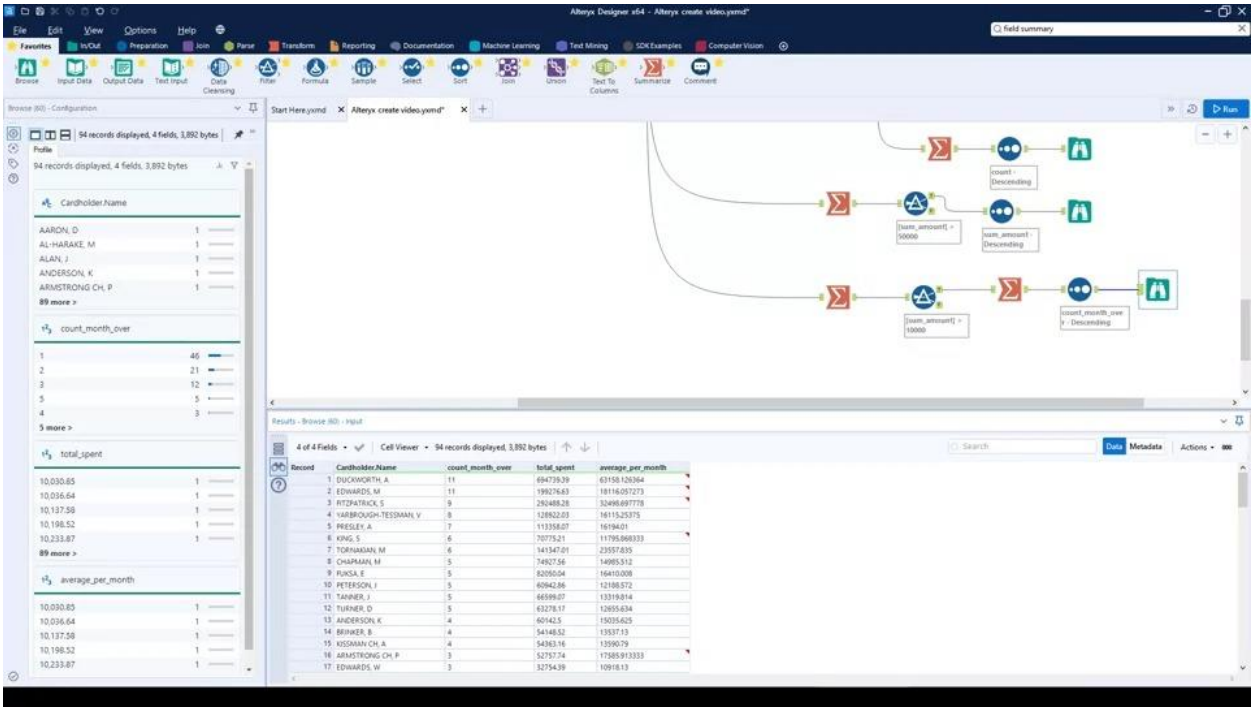
There we are, and let's go over here. The first thing we're going to do then is we summarize is to group by cardholder name because this is an analysis by cardholder, by a person and so a group by cardholder name. The next thing that we want to do is sum up the amount because this is by year. We have one year of data, so we're getting a total for the year and we want to make sure it's not over $50,000. We click on amount here and then we're going to sum that up. Again, we'll make this lowercase. Then the next thing we want to do is count those transactions and make sure it's on amount. We'll just count. We're going to count and see how many times, how many transactions they used to violate the amounts. That's done. The next thing we need to do though, is we're only going to look at the totals that are above $50,000. We need a filter, we need to get only the true, only the T for those that are above, where sum of amount is above $50,000. That's what we'll do next. We'll just have greater than 50,000 there. Next, let's sort. Again, we're going off of true here, and we're just going to sort by the sum of the amount and let's look at the biggest ones on top and then let's go ahead and take a look at what that does. Let's run this. What do we learn from this view of the data? Well, actually we learn quite a bit. We see that there are over 60 violations. That's 68 individuals that have spent



Record	Cardholder.Name	sum_amount	count
1	DUCKWORTH, A	700494.76	2662
2	MATTHEWS, W	498893.62	2522
3	FITZPATRICK, S	313602.17	267
4	EDWARDS, M	206108.18	498
5	YARBROUGH-TESSMAN, V	161540.47	223
6	REUTER, D	157500	833
7	TORNAKIAN, M	156848.03	655
8	PRESLEY, A	138183.51	156
9	CHAPMAN, M	117984.01	353
10	TURNER, D	117205.13	471
11	TANNER, J	110928.47	235
12	HOLMES, K	107442.03	27
13	LIANG, Y	106265.67	197
14	ANDERSON, K	106176.67	347
15	FUKSA, E	105780.63	153
16	PETERSON, J	101654.5	380
17	KING, S	100570.59	201

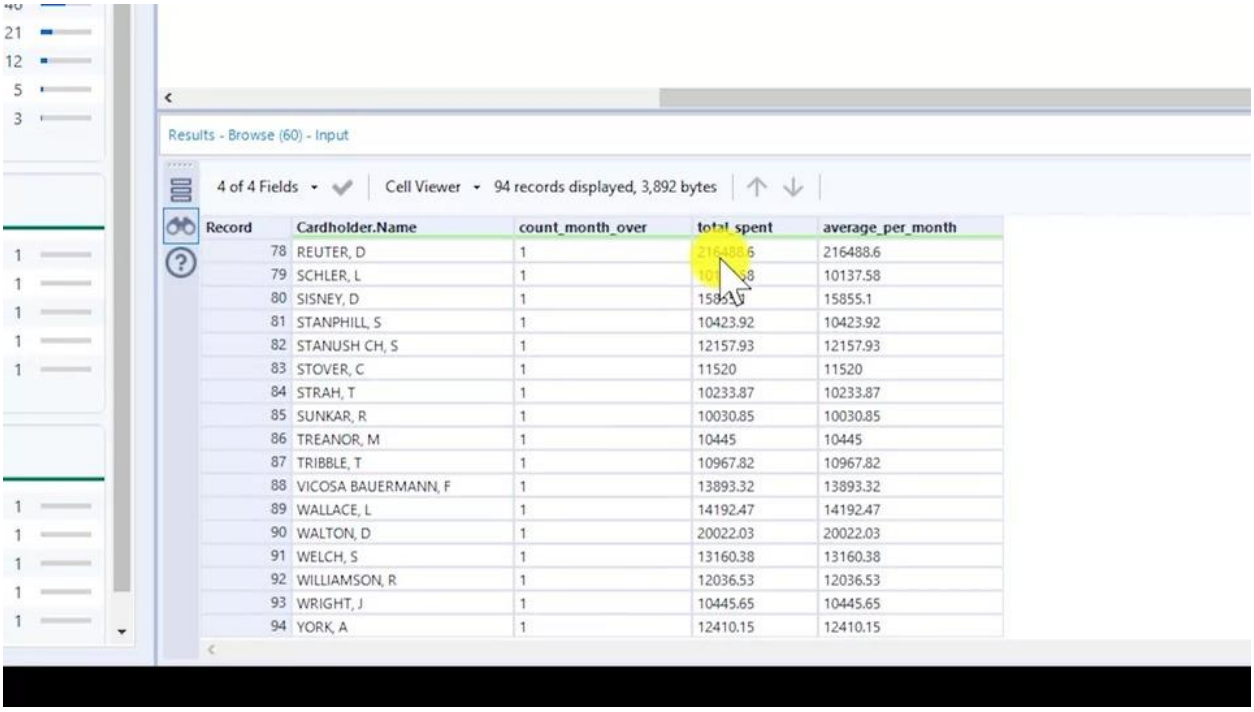
more than the $50,000 limit in a year. This is a big number. It's like three percent of all cardholders for the year. Some of the three percent of the people and that's something that the order needs to be concerned with. It is certainly possible at all of these or at least most of these transactions were authorized and appropriate, but the order clearly needs a better understanding of what this control really should be if these many people are violating it. We also learned that some individuals have exceedingly high numbers of transactions. In fact, we have two, we have Duckworth A and Mathews W and these individuals have more than 2,500 transactions in the course of the year and you can see these two are big outliers relative to everybody else. They have a lot more than everyone else. They incurred together more than a million dollars of expenses. Clearly these two individuals need to be investigated more fully. Again, it is perfectly possible that these individuals were authorized to make all of these transactions, but they clearly have roles that differ from the roles of everyone else in the organization, since they're creating thousands of more transactions than everybody else, the auditor definitely needs to sit down with both of these individuals, most likely you have to understand what their roles are in the organization and how they use their purchase cards. Next, the order will, of course, need to verify what they say through appropriate audit tests. That is the auditor trusts but always verifies. The auditor must remain skeptical all times and can't just trust what management tells them without testing for themselves. Finally, if indeed these individuals are authorized to make all of these transactions, the auditor should look at compensating controls that help make sure that these individuals do not do anything inappropriate. That is, if they're allowed to spend significantly more than the $50,000 limit, there need to be other compensating controls in place to prevent these two individuals from doing something inappropriate. Finally, we have one more control

to test. That is, we need to examine monthly totals to ensure that individuals are not spending more than $10,000 a month. I think that we all know by now that we're going to find some violations here so let's dig in and see what the data tells.



Let's generate a report that's slightly more complicated than the others we've looked at so far. That is what we would like to generate a report that lists by individual how many months violated the $10,000 spending limit, how much total was spent in those violating months, and the average amount that was spent in those violating months. Thus, to generate this report, we'll need to first summarize by individual and month. Next, we'll need to filter by 10,000 and then we'll need to summarize again, grouping by individual. Let's start out with our summarize here. We'll bring that down. Give ourselves a little bit more room. Pull the data in. This is our first group by cardholder name and so group by cardholder name. We then need to group by month. It this is a little more complicated because we're not looking by cardholder over the whole year. This time we're looking by cardholder, by month, so let's go to month and do that. Our second group by, and then we just need to sum up amounts because we're going to filter based on whether it's above 10,000, so we sum. Again we'll case this. Next filter. In the next thing we're going to do is filter to make sure that to look at only the ones that are above 10,000, so we'll click on the "sum_amount" greater than 10,000 here. Then next, we'll pull in and then I'll summarize. Again, the goal is here once we have the violating transactions, so we have every time there's a violation of 10,000 for a month. We'll have multiple violations potentially by individual. We want to group those down and just look at one observation per individual and then get some summary statistics about those transactions. Here we're going to group by cardholder name because we want to get one observation by card holder. Then let's just do some things on the amount, so let's count to see how

many of those transactions for that month or more violations and then let's go ahead and look at the sum of the amount. What's next? Sum these up. Then let's finally look at an average. Again, the average is down here under numeric. Again, what we have here is just like a count of the months that were over. Then here we have a total span. Then here we have the average per month. Finally, we'll sort that. Let's just sort it descending by the account that there's again multiple ways to do this. Well, let's go ahead and get counted months over and insert that by descending. Then we'll take a look at it. Let's run that. Verify that we don't get any errors and we do not and we have 94 observations, your 94 violations, so let's take a look there.

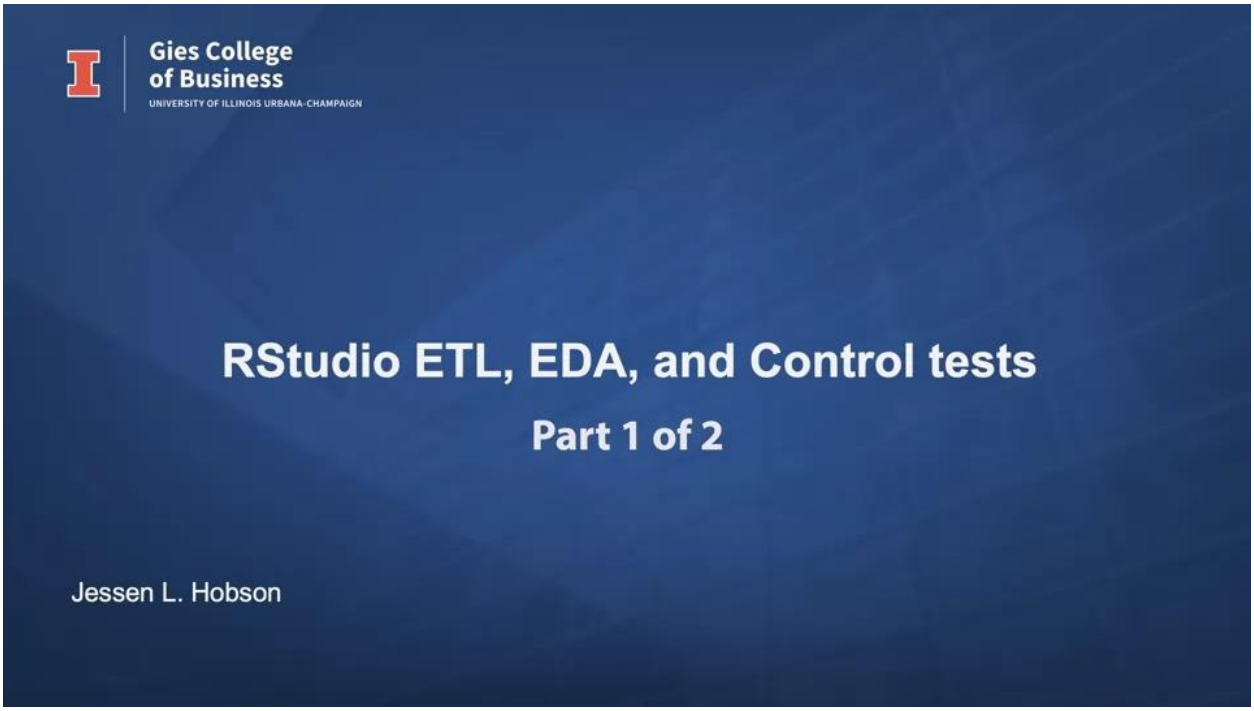


Record	Cardholder.Name	count_month_over	total_spent	average_per_month
78	REUTER, D	1	216488.6	216488.6
79	SCHLER, L	1	10137.58	10137.58
80	SISNEY, D	1	15855.1	15855.1
81	STANPHILL, S	1	10423.92	10423.92
82	STANUSH CH, S	1	12157.93	12157.93
83	STOVER, C	1	11520	11520
84	STRAH, T	1	10233.87	10233.87
85	SUNKAR, R	1	10030.85	10030.85
86	TREANOR, M	1	10445	10445
87	TRIBBLE, T	1	10967.82	10967.82
88	VICOSA BAUERMANN, F	1	13893.32	13893.32
89	WALLACE, L	1	14192.47	14192.47
90	WALTON, D	1	20022.03	20022.03
91	WELCH, S	1	13160.38	13160.38
92	WILLIAMSON, R	1	12036.53	12036.53
93	WRIGHT, J	1	10445.65	10445.65
94	YORK, A	1	12410.15	12410.15

What do we learn from this view of the data? This report shows us that over 90 individuals violated the monthly transaction limit of $10,000. About half of those violated the control for more than one month. Thus, about 5 percent of the individuals that hold purchase cars violated this control. Now, as before, the order will need to form an understanding of why this is the case and then test that understanding. Also as before, there are a few individuals that stand out for special investigation. For example, we have Duckworth a this individual had 11 different months that violated the $10,000 control, so all but one month in total, Duckworth spent nearly $700,000 in those violating months and an average of 63,000 of spending in each of those violating months. This is obviously a significant outlier that the auditor needs to examine. Now, Edwards, M also violated the control 11 out of the 12 months. However, these violations were not quite as large as those from Duckworth since the average violation was $18,000 only. Let's take some additional looks at this. We were to sort based on the total amount spent and the average spent per month. I won't do that, but we can just look at it. We can see that there are other outliers as well. For example, well, Duckworth has the highest total

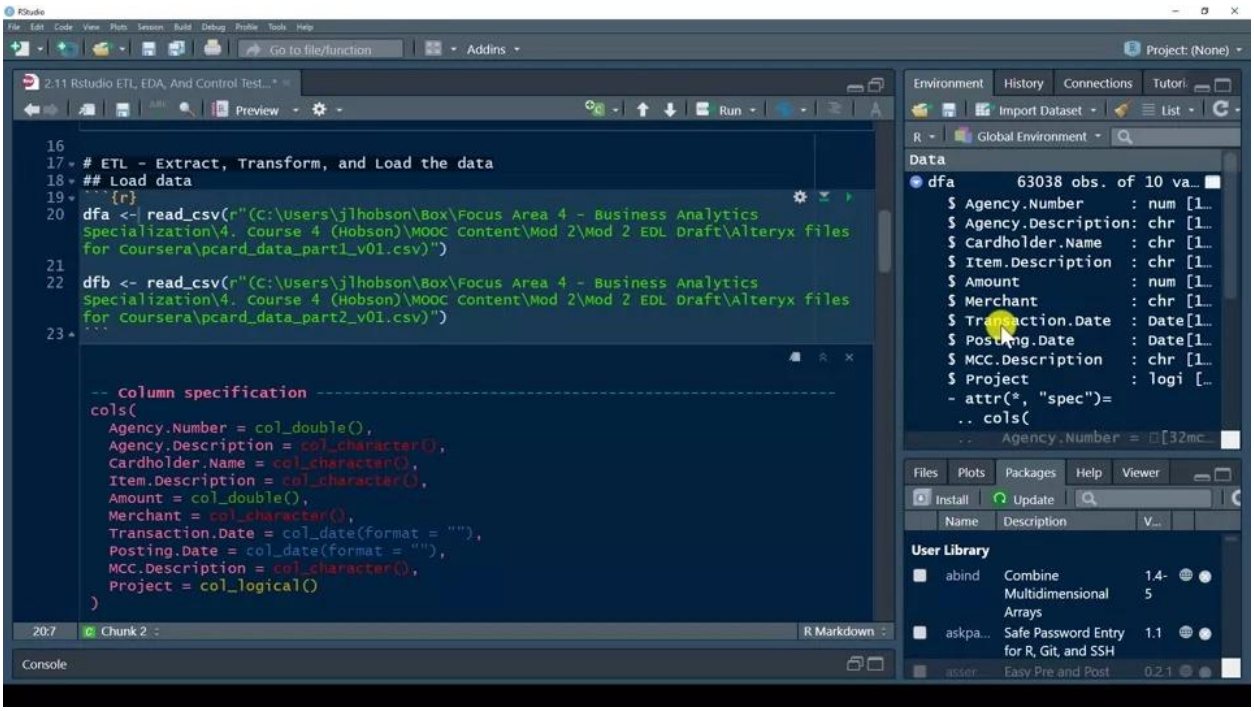
violation spinning. Matthews, right here in row 23, also has $620,000 a violation spending. You can see right there, this is very curious because this happened in only three months, that is Matthews average spending in those three months was more than $200,000. Which clearly needs some significant investigation by the order. Next, if we sort by average violating amounts, again, we won't do that. But we can just look at it. We can see that routers D, let's see right here in row 78 at the top, only violated the control in one month. But that violation was for a whopping $216,000. Thus, on all there is clearly a lot of work for the odder still to do.

Lesson 2-12: Screen Shot Solutions in RStudio - ETL, EDA, Control Tests Part 1

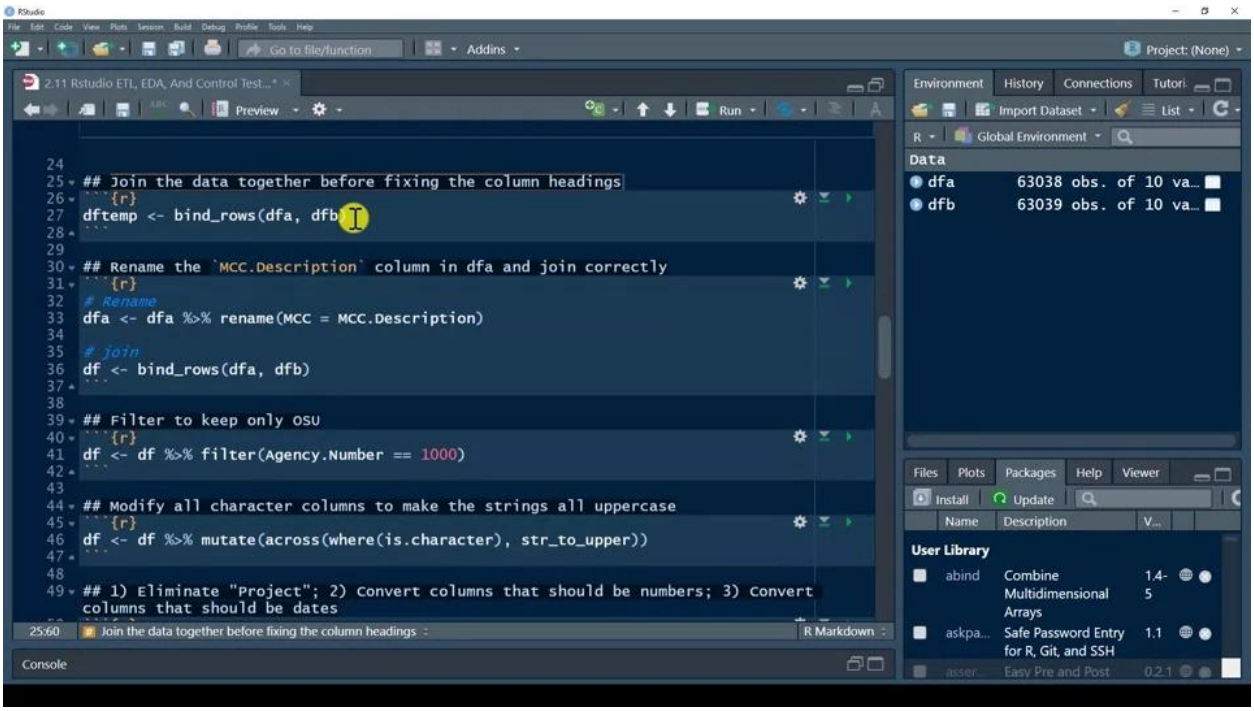


[Music]

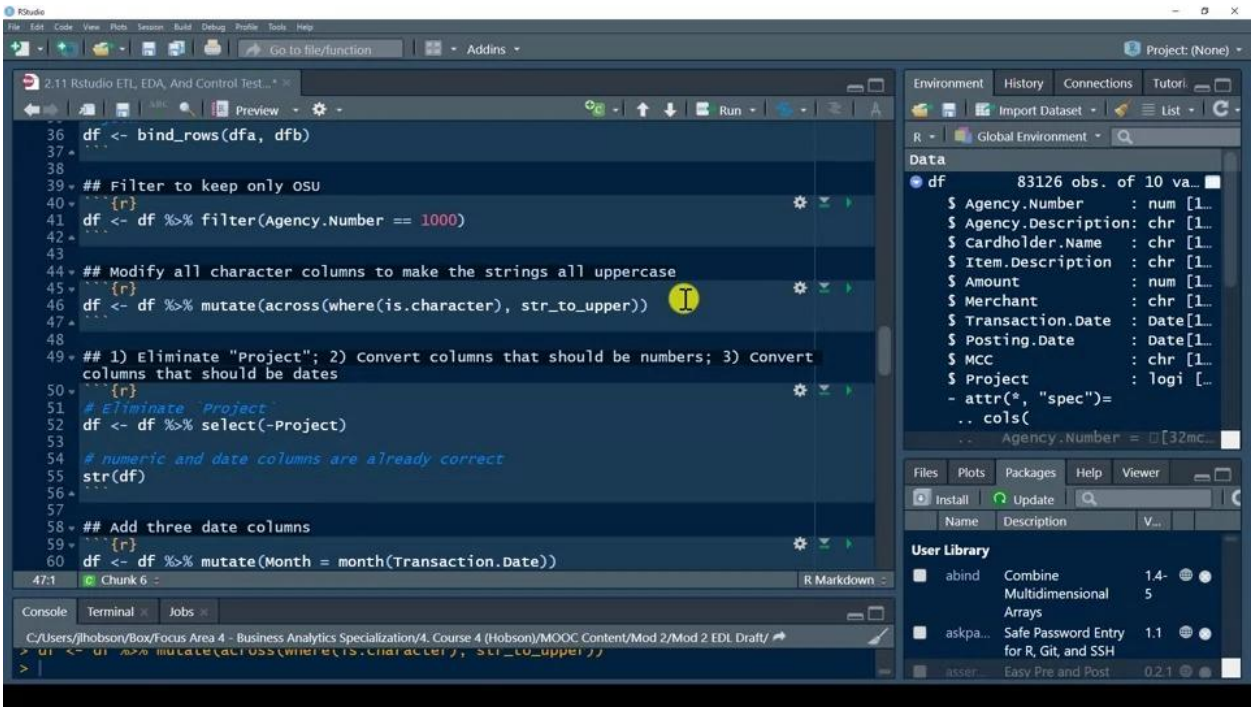
This video replicates analysis done in all tricks for ETL, ED A, and control testing for the procurement cards from Oklahoma State University. The goal of providing this video is to give you a comparison of the two tools. This will help you gain proficiency in both tools, both all tricks and our studio and will help you see which of these tools you'd like better and are more likely to use in your future. Now, if you need a refresher on your our studio skills, please see the refresh your video about our studio in this module. Please also consider reviewing other introductory videos about R and R studio from prior classes. As always, I strongly encourage you to actively participate in these videos by doing everything that I'm doing, load the data into our studio. Follow along with each line of code by typing each line out yourself instead of just copying and pasting. This is the only way that I've found to successfully learn and be proficient at these skills.



So let's go ahead and start our studio and so we're going to load up this our studio notebook and go through and do the same things that we did in all tricks. So as closely as we can using R. This will give you the opportunity to compare and contrast R versus all tricks, see which one you like better, which one works better for you and help you grow your toolbox in that way. So let's load the packages that we're going to need. We're going to use tidy verse and lubricate these of course have already been installed by me and you can see they come up. And we won't use every single package in this meta package of of tidy verse but we'll use some of those. Now if you don't have these packages you'll have to install them right? So you go over here and install but I have them and next we'll load up the data. So here's the data and I've got both data sets. I'm going to call this first one and DFA. And I'm going to call the second one DFB and then this is just the folder, the path, yours will obviously be different. The R here is going to allow this to be a real string and so it's not going to make me escape these weird symbols and it's just going to bring them in as they should as I mean them to come in. So I'm going to run this and load this data. So here we have on the right we have DFA, DFB 63,038, 63,039, 10 variables each. This is exactly what we had in all tricks.

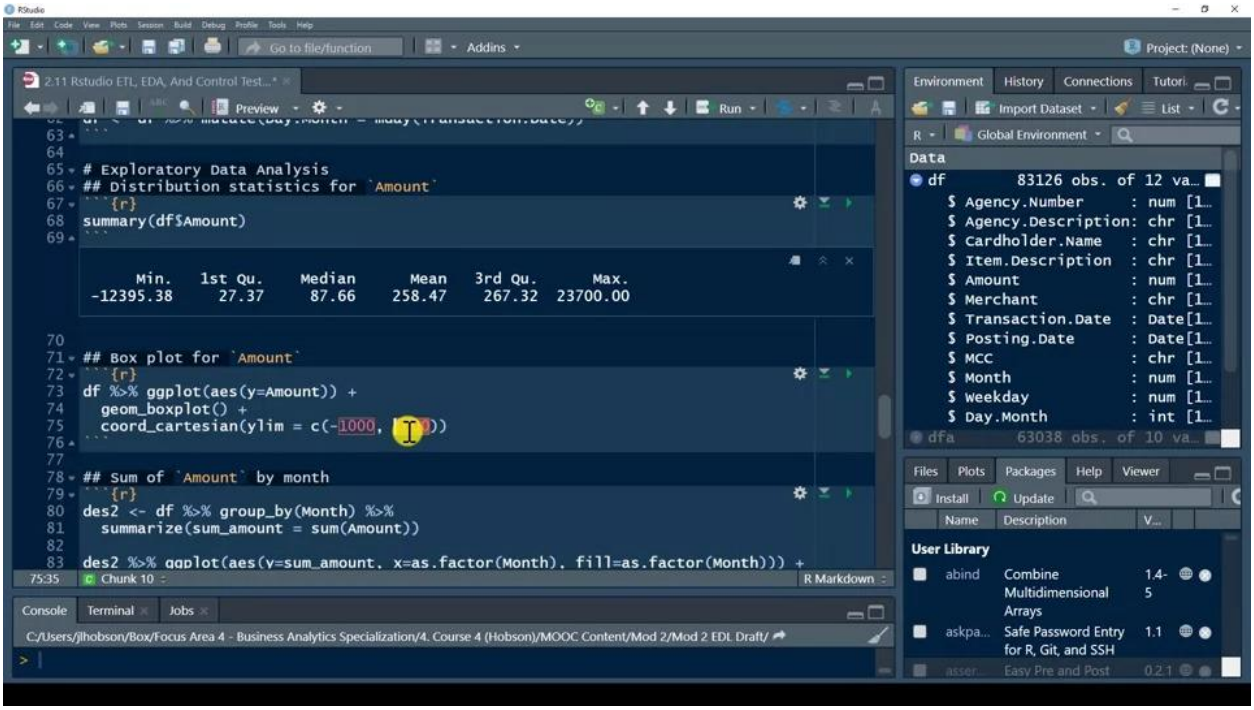


So great let's go down and do the next step. We had an all tricks which has joined the data together. But before fixing the columns. Remember the first data set had MC description instead of MCC. And so again we're going to try and do is stack these rows together and we can't do that when we have one of the columns with a different name. We need the columns to have the same names exactly.



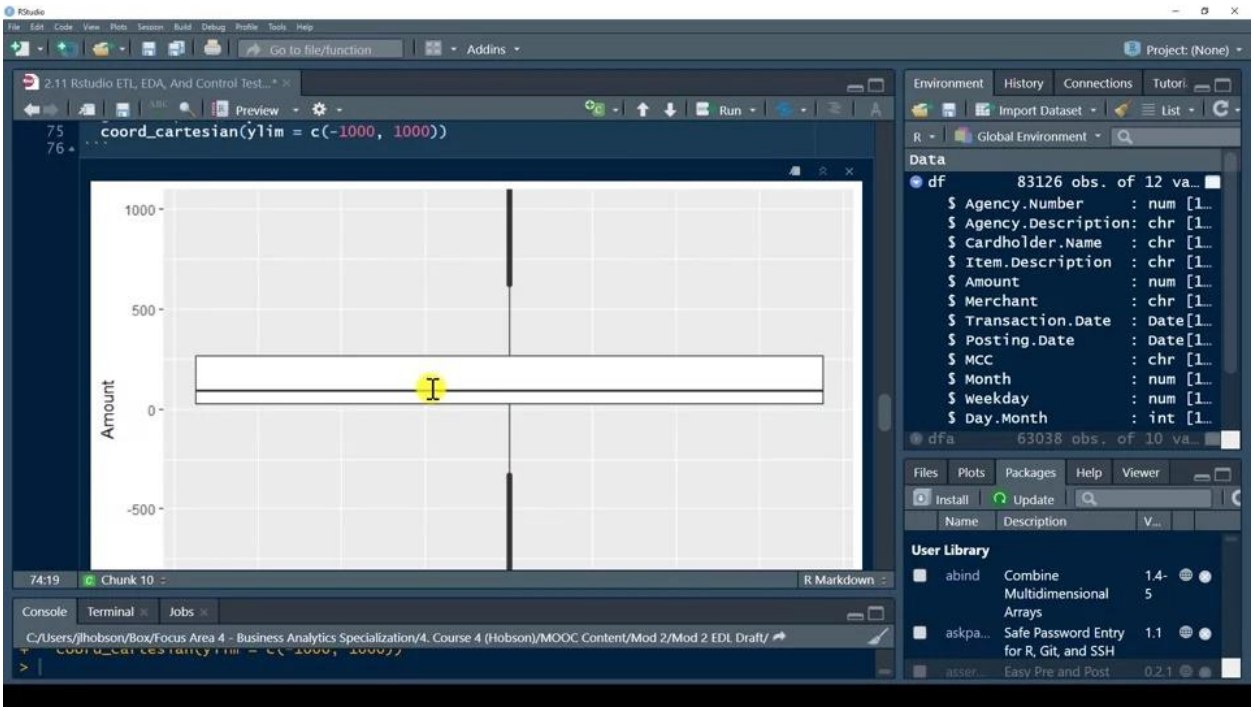
So, we'll go ahead and run this. Again, this bind rows just binds the rows. It stacks the rows together from the first data set and the second data set into this DF temp data set. And as we look to the right, we can see this was not done correctly right? Because we have although we have the right number of rows 126,077 we have 11 columns. We have both the MCC column and the MCC description column, which is not what we wanted. So, what we need to do then is first to rename MCC description to MCC and then join these up together. So to rename them, we'll use the rename function, which is a handy function. We put the new name that we want in this case MCC. And then we put the old name of the column so it knows what column to change, all right? And we'll create this in DFA and then replace the old DFA with that new function. So that is done. We can double check it if you like. MCC as we want it, excellent. Next, we'll bind the rows together as before. All right, so DF is our new data frame which has everything in it. We've got the right number of rows and the right number of Columns. We've got no MCC dot description just MCC. So that worked as expected. All right, our next step is to filter and only keep the rows that are from Oklahoma State University. Require that the rows, that's our Oklahoma State University have agency number equal to 1000. So let's use the filter function and we'll use filter and then again, we'll just have agency number equal to 1000. We'll create we'll do that in data frame called DF and will replace the old DF with that okay? So that's done and you can check it if you like. We can double click here and take a view and see that it looks like it's all just Oklahoma State as we wanted. Okay, so that's great. Next let's modify all the character columns to make the strings uppercase. Now, this takes a bit of explaining because what we're going to do is have our help us do this and what we're going to have it do is using three nested functions. We're going to use the mutate function to look across all of the columns and then pick

out the columns that are characters already and then change the text to uppercase. So we can do that without having to write down all the columns because when the data came in it did so in a really smart way. It had the columns that should be numbers like agency number and amount come in as numbers. And it had the columns that should be character should be strings come in as strings such as agency descriptions such as MCC and and so on. Finally it had the the transaction date and posting date which should be dates coming as dates. Now our was able to do this because what it does is it guesses the column. The format the columns should be based on the first few observations in that column. Because our data is fairly clean in those columns had the first few observations were in the right format, it brought everything in correctly. So all we have to do again is tell the mutate function to look across and find every column that is a character and change the text, the string in there, to uppercase.



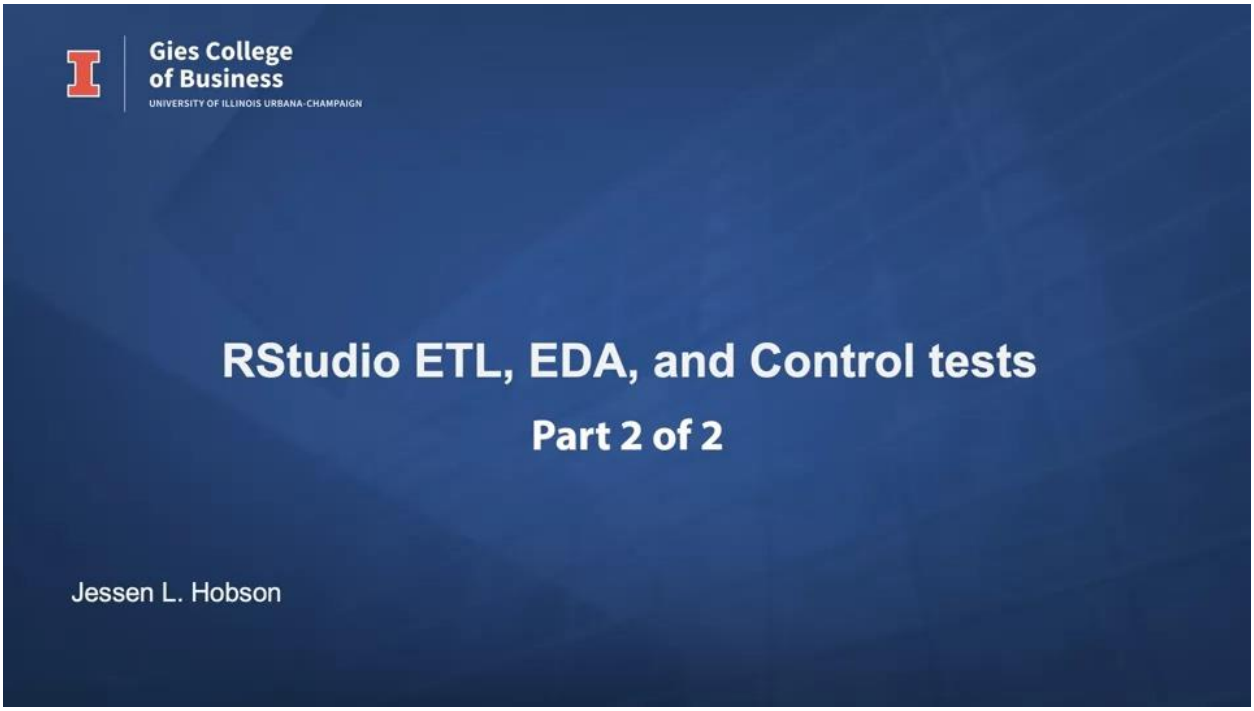
So we do that in the data frame and let's just look at this briefly, everything looks now to be an upper case as we wanted. So that's that's worked. So that's excellent. Our next three steps let's eliminate that project column because it's no good. Let's next convert columns that should be numbers to numbers and convert columns that should be dates to dates. Which again we're not going to have to do right as we just saw they came in correctly. So the only thing we have to do is delete project and we do that using the select function by just saying not project. So this minus in front of project says don't bring that column in to this new data frame, which will this new data frame, which we'll call DF. Which replaces the old DF. So we do that sure enough, we have 9 variables and we do not have the project column anymore. And then we can just look at this if we want, have scrunched it up here to make it bigger and easier to see. So this doesn't look very pretty, but things have worked as we expected. Our next objective is to create 3

new columns based upon the transaction date column. Again, this is just replicating what we did in all tricks and we'll use the handy month w day an m day of functions that come from lubricate. So this is from the lubricate package. And what we do here, we also use mutate. So again, mutated just a tidy verse function that helps us take and create new variables and change variables. So we'll use the month function to create the day of the month using transaction date and we'll call that month as we did in all tricks. And again, we use mutate, we put that in our our new data frame called DF and replace the old data frame DF. We'll use w day and then we use m day, alright to again replicate the results from all tricks. We can look down and see that we have month, date, day, month. Let's look at this real quick just to make sure it kind of worked that was helpful for us and all tricks. So as we go across we have transaction date. All right so this should be nine for the day of the month. For the day of the week it should be to which we would necessarily know, but that was the same in all tricks. And for the day of the month it should be the 28th. So it looks you know like this worked correctly. Okay so great. What is our next step? Our next step then is to start the exploration of the data. So we're done with the ETL, let's go to EDA. So the first thing we did is we kind of looked at the distribution distribution statistics for amount. And so the summary statistics were helpful for us. We can do this with the summary function by and just look at this column from DF. So we indicate the column amount with the dollar sign here. And so we have as we did in all tricks, a negative minimum that we noticed before. A really high maximum, we noticed that before and the mean is higher than the median. The next thing we did was a box plot four amount. Let's do that, and here we'll use GG plots,all right? So GG plot is a really great function from tidy verse. It is super flexible but sometimes confusing. The way I think of it as the layers of a cake, so the first thing when you do is called GG plot, all right? And then we're going to put in aesthetics, that's what AS mean. It's not a great name by aesthetics, just going to indicate different parts of what the graph is going to be, what the plot is going to be. In our case, we're looking at one variable and that variable is amount and that's going to be on the Y axis, right? And so we're just going to say why equals amount. The next layer of our cake is going to be box plot. So, Gm underscore box plot just means, hey, let's do a box plot. And then finally this chord Cartesian, our third layer of the case is just going to say, let's make Y axis be constrained to be within negative 1000 and positive 1000.

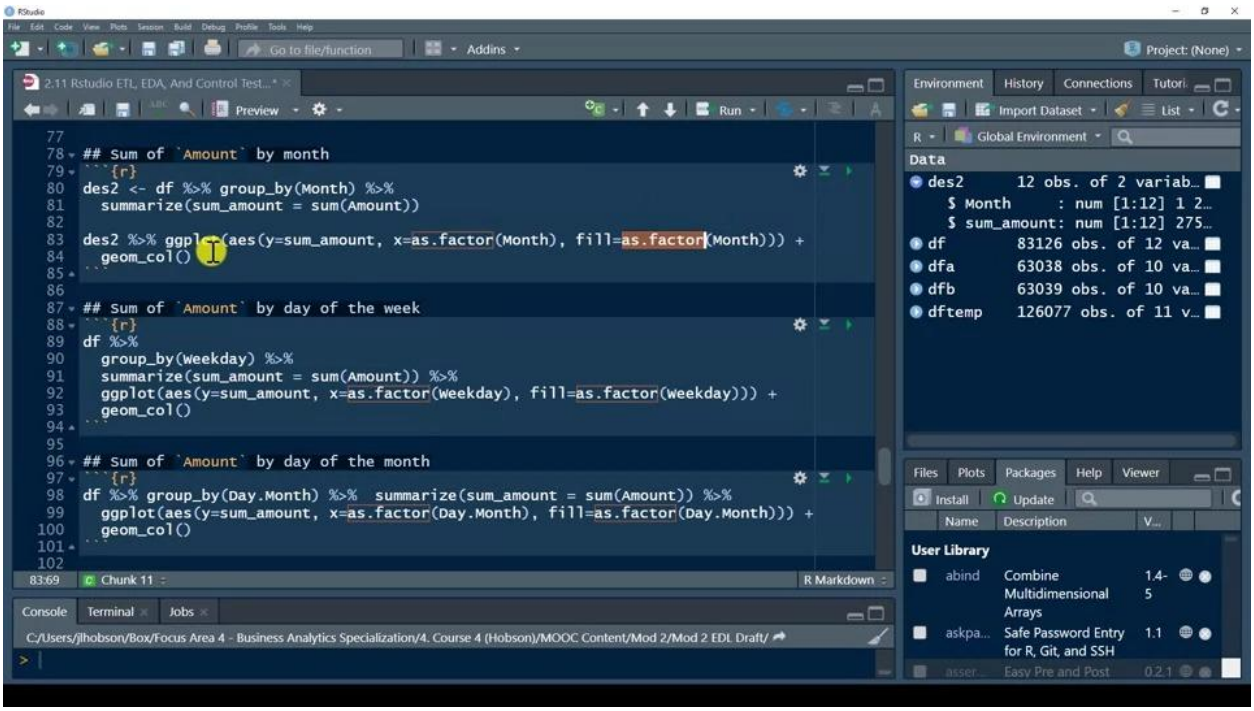


Again, this is just the same thing that we did in our all tricks. And you can see we get a very similar looking graph.

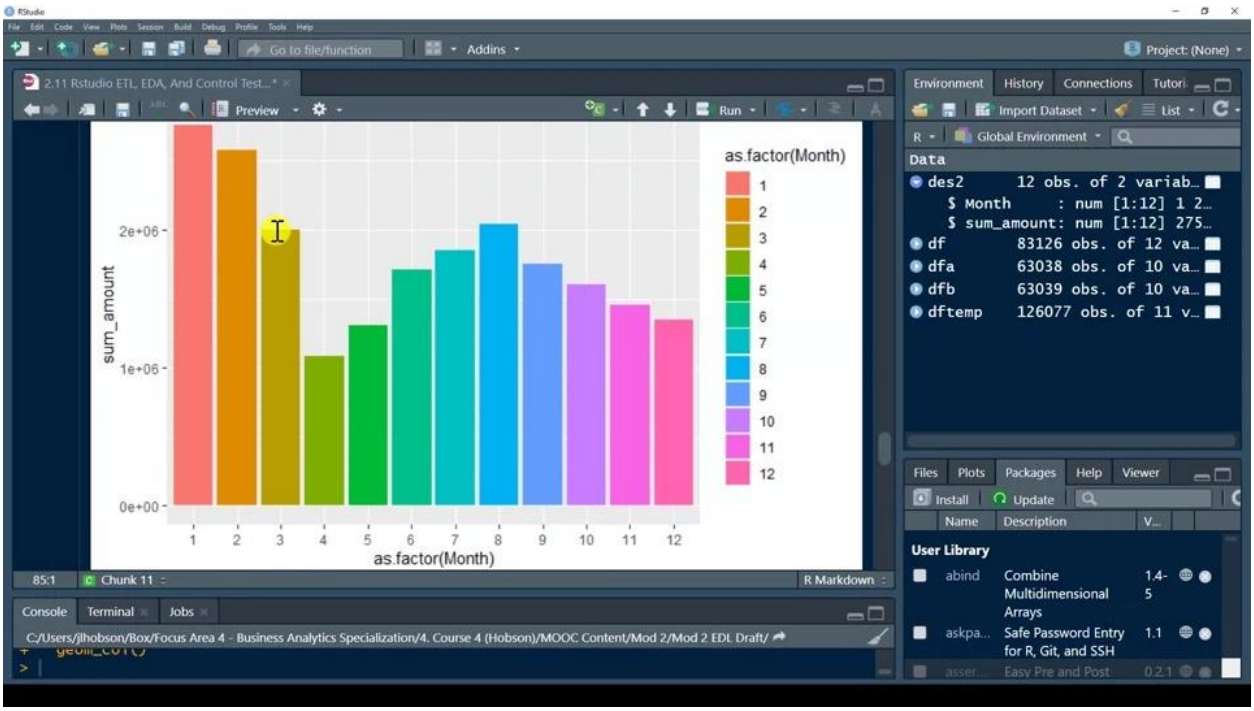
Lesson 2-13: Screen Shot Solutions in RStudio - ETL, EDA, Control Tests Part 2



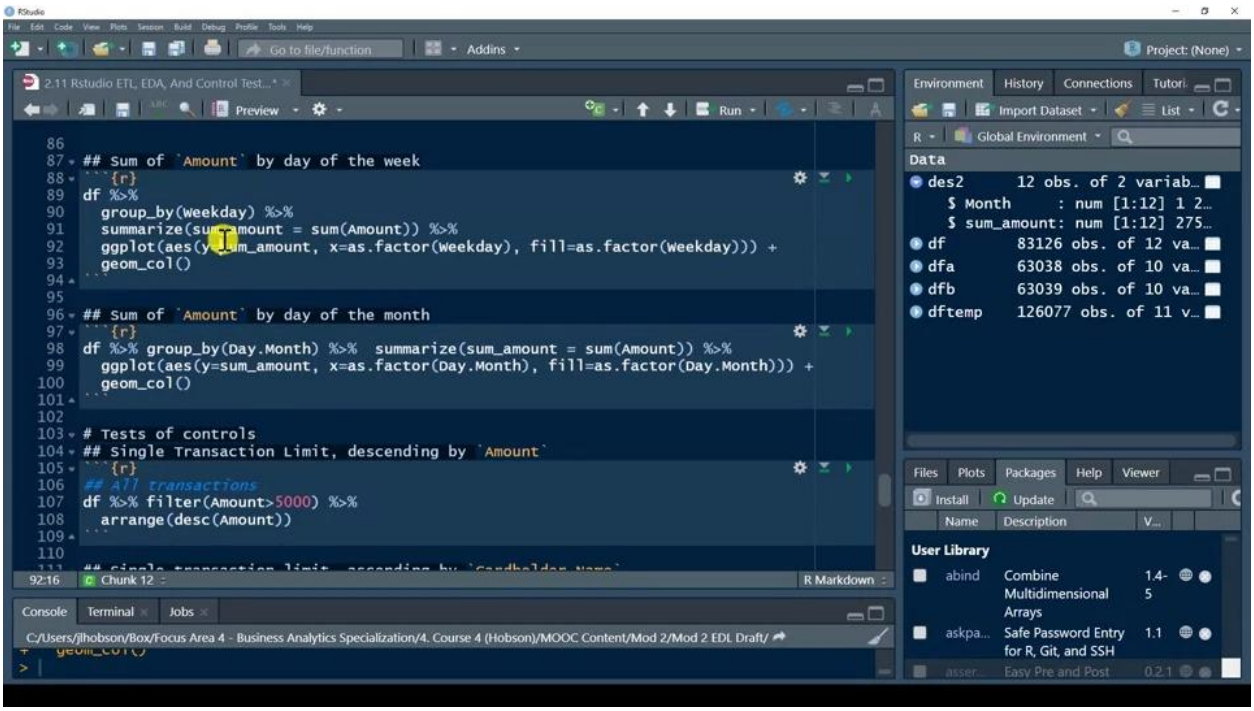
[Music]



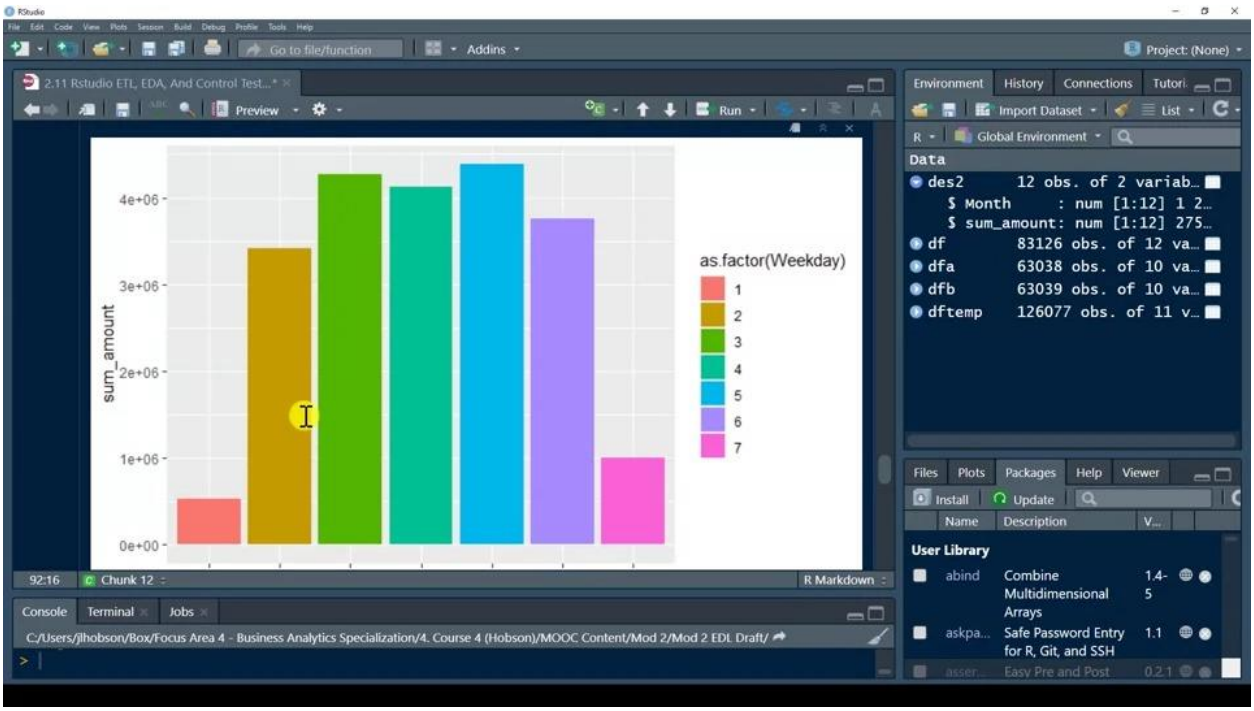
Excellent. Let's go then and do our next step here, which is we're going to sum amount by month. We want to get the sum by month, and then graph that. Let's go ahead and do this in parts. First, we need to use the group_by as we did in Alteryx, and summarize. We'll call this des2, so description 2, and we'll take our dataframe, and then we want to a group by month first, and then we can pipe over, summarize. Then we'll take our new column, which will be called sum_amount, as it was in Alteryx, and sum amount. Let's just run that, and we can go look at it. Again, recall what we did before in Alteryx, is we simply have one observation per month, which is the sum of all transactions within that month. Next we'll graph that. We take des2, which we just made, and then we're going to have our ggplot layers. Here's ggplot and our aesthetics now. We're going to have a column chart in ggplot. Here we have the aesthetics. On the y-axis, we'll have sum of amount. We want to have each bar raised as high as the sum of the amount for that month. On the x-axis, we have each month, and month came in, as you can see over here, a number. But we want to make it discrete, we want to have it be a factor. We don't want it to be a number, we want it to be a factor, so we just add this function right here, this as.factor function. Finally, we want to fill in each different column in a different color by month. Each month is a different color, and so the last piece of our aesthetic is fill, and we'll have equals month, and again we'll add the as.factor. The second layer to our ggplot cake is just that we tell it what plot it is, and it's the column plot.



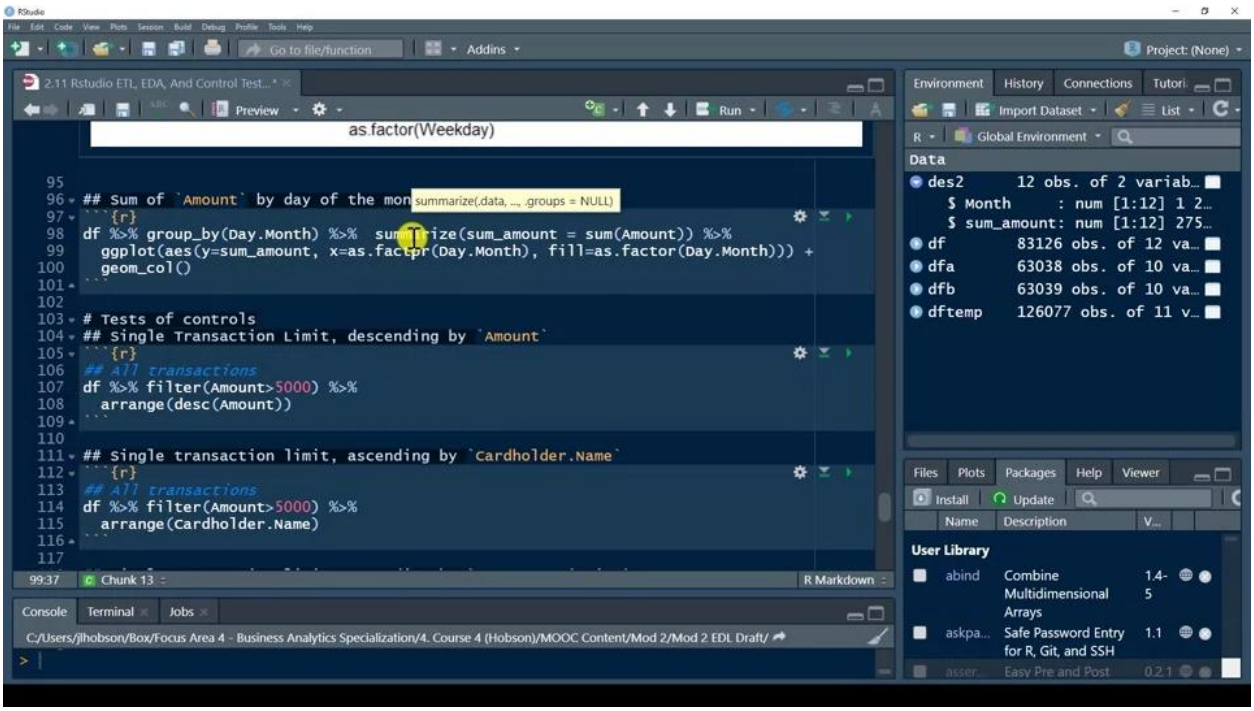
Let's go ahead and run that and take a look. You'll see this seems very familiar. It looks very much like it did in Alteryx, where we have the greatest number of purchases in January and February, we have another high amount in August.



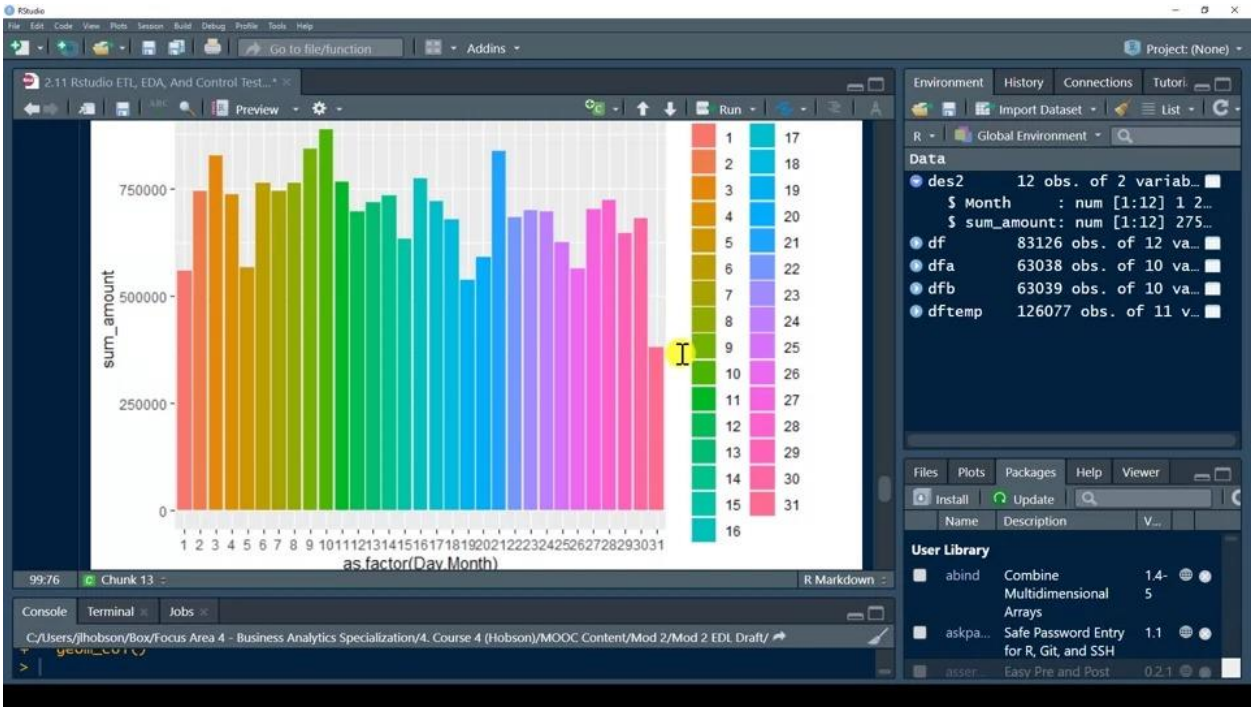
Next, we're going to do a similar thing, we're going to get the sum of amount by day of the week. This time I'll just do it all in one step. Same thing. I could do it in two steps, I'll just do it all in one, where I'm first going to group by weekday, so I want the day of the week to be one row in this new group_by dataframe here. Then I'm going to summarize amount again as I did before. Summarize, I'm just going to take the sum of amount and create this new sum_amount column. Then I'll just do ggplot, I'm going to do another column plot. It's going to be very, in fact exactly the same here. In fact I cut and paste this, and we're just going to have weekday and weekday instead of month.



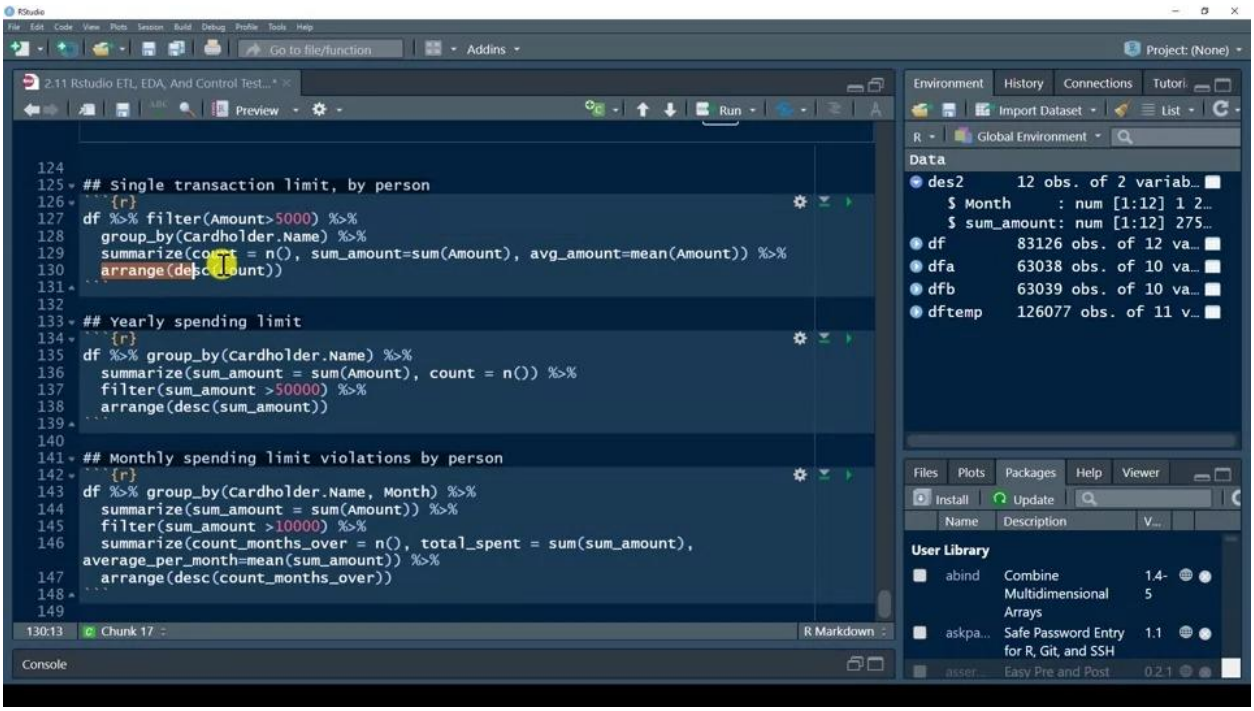
Let's run this all together and take a look. Again, it looks very similar to what we saw in Alteryx. We have most of the transactions happening during Monday, Tuesday, Wednesday, Thursday, Friday, but there is some months on Saturday, there is some on Sunday.



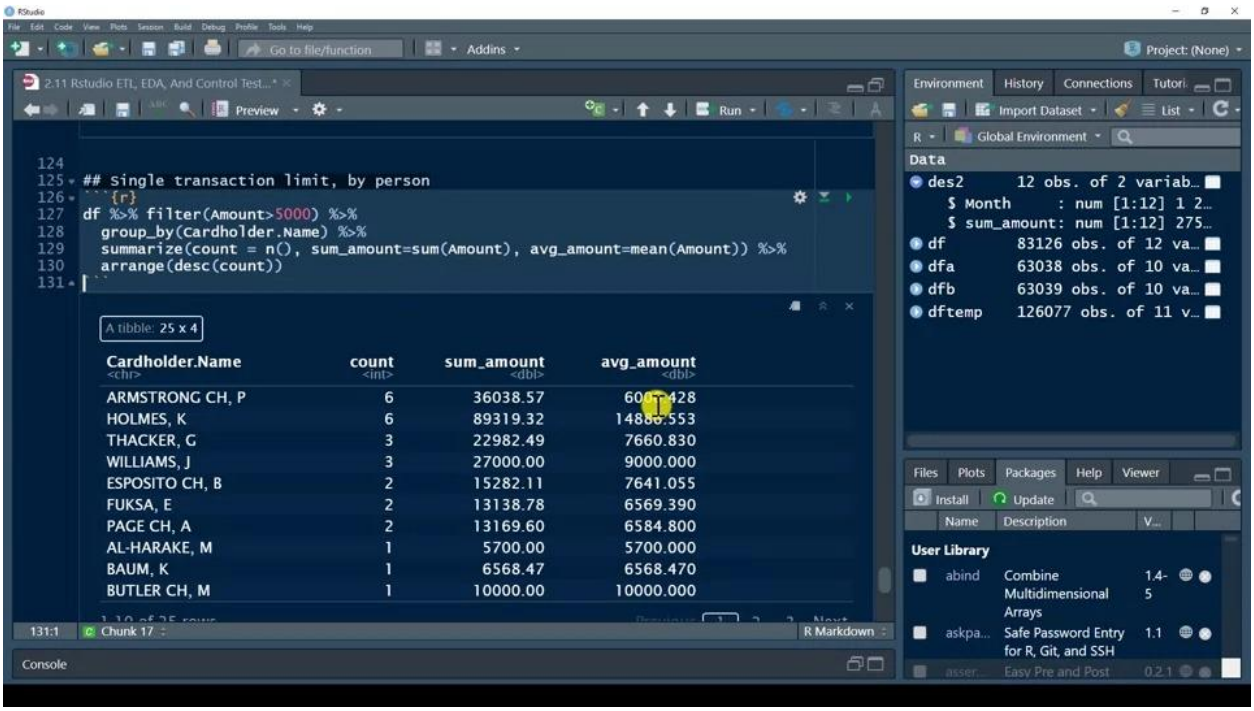
Finally, we'll do the last piece of this, which we're going to group_by the day of the month. That's what this group_by is. We summarize and sum amount again, and then there's our ggplot. Just really cut and paste from before.



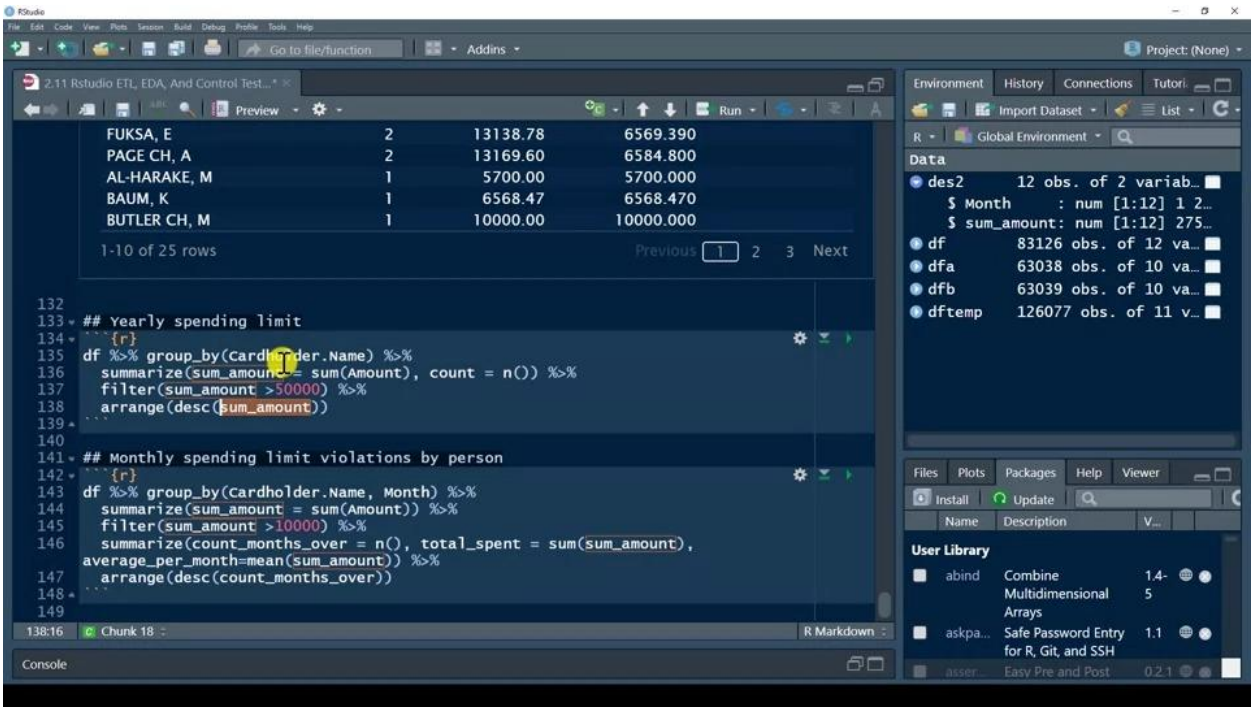
We just change to day.month in here, and there we have it. We can see we didn't learn a lot from this graph, although you can see that not very much happens at the end of the month. But there's quite a bit of variance here.



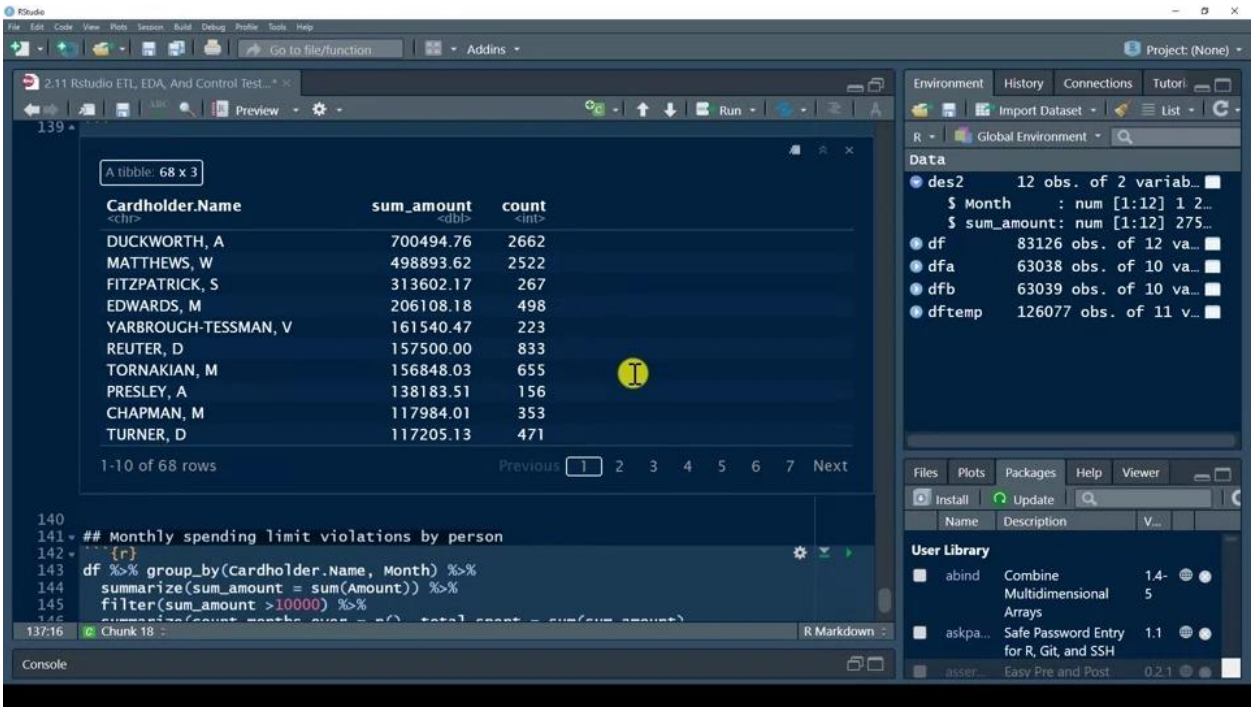
Our next step is to test our controls. We're going to again just follow right along what we did with Qualtrics. We started in Qualtrics by looking at the control which said that an individual transaction was not supposed to be over $5,000. The first thing we did was we looked at three different views of all of the transactions over $5,000. Here's the first one. We're just going to filter and pick out the transactions that are greater than $5,000, and then just sort. We arrange descending by amount. We nest the function in here, which is descending by amount. Let's run that, and it shows up here. We can scroll across and look at what we have. We see our buddy, homes shows up again with that big 23,700 transaction. That works just like it did in Alteryx. Next, we're going to sort or arrange by cardholder name. We just change this here, and we'll do it alphabetically, so we don't need to add that descend function. Here we have, again, the same analysis that we saw before. In Alteryx. One more time. Let's now sort by item description. Again, same thing. We had the 600 plastic shields we saw before, the air travel, the general purchases, which were very leery of the room charges, and so on. We're going to do one more view then of this individual transaction control that we're not supposed to have a transaction greater than $5,000. This time we want to look at it by cardholder name again, just what we did in all tricks. We're going to group by cardholder name after we filter and look at only the transactions above $5,000. We're going to summarize now, add three columns. We want to add the count. We want to add the sum of the amount by individual for only the transactions that violated, that were above $5,000, and look at the average of those transactions. Finally, we'll sort by count descending,



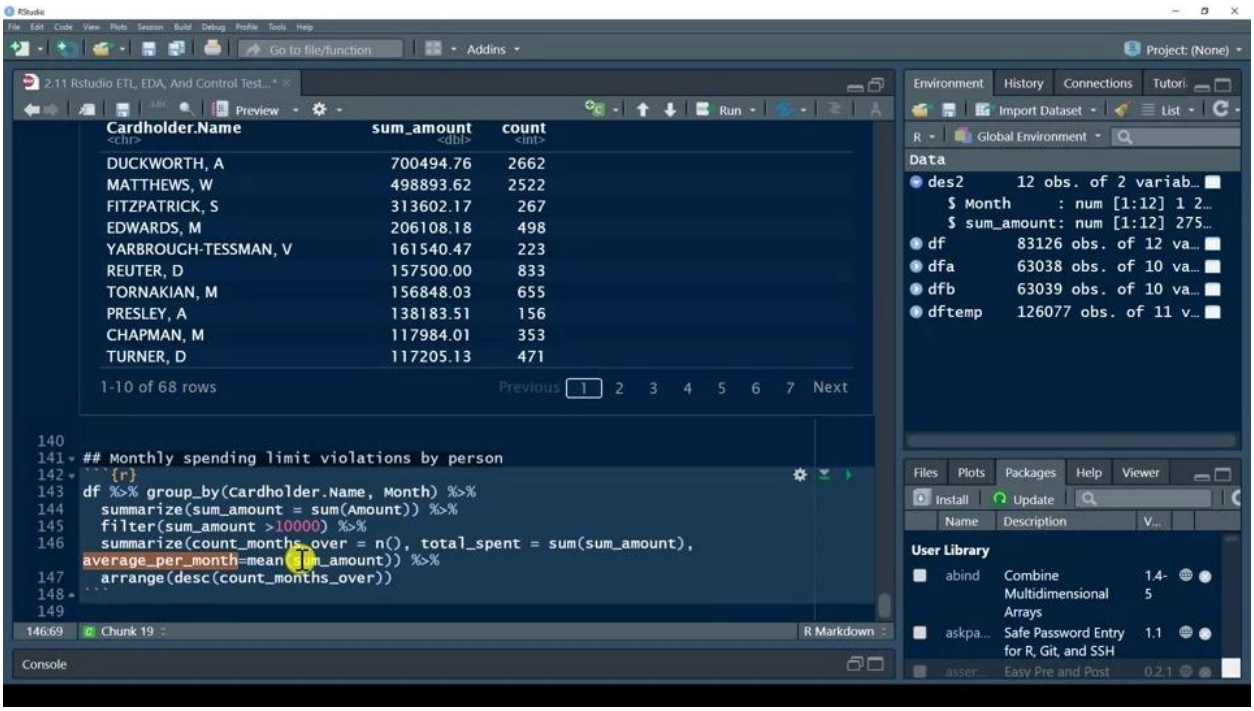
and so let's run that. We have again the same analysis we had an Alteryx where we have Armstrong with six for a total amount of violating transaction. These are just the transactions that violated $36,038 and the average $6,000. Then Holmes K, which is really worse, where we have six different violating transactions for a total and those six totaled to $89,000, but even more problematic, that average of those six transactions was almost $15,000, which is three times the transaction amount that it shouldn't be. It's three times $5,000, and so these are way high, something definitely the auditor needs to look into.



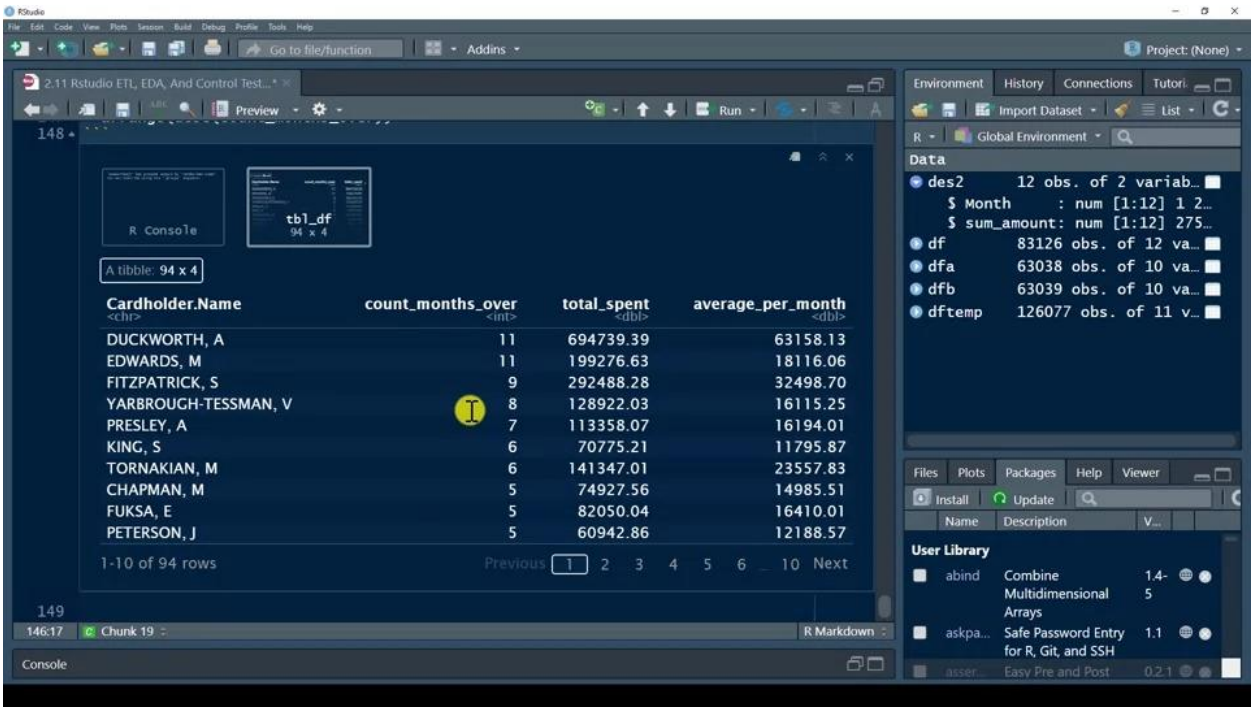
Let's now move on to our second control, and that is that the sum of all the transactions for the whole year for each individual should not be over $50,000. Each individual should not have a total of transactions that goes above $50,000. We'll do the same analysis we did in Alteryx. Here we're going to group by cardholder name because it's by individual over the course of the year, and we only have a year's worth of transactions, so we can just group by cardholder name. We then summarize, and we're going to sum the amount, and we're also going to count. Then after that, we'll filter by greater than $50,000, and then we are going to sort by sum of the amounts.



Let's go ahead and run that. We again see we have two outliers. We have Duckworth and Matthews, which have a sum of their total transactions rather of $700,000 and $498,000. Over a million dollars worth in just one year for these two people put together, and they have over 2,500 transactions each. These two people are doing something different. It might be fine. The auditor's going to need to have to go check that.



Last but not least, we'll look at our final control, and that's the monthly spending limit that, monthly, an individual's should not spend over $10,000. This is a little more complicated. We have first to group by cardholder name and by month. This is going to give us one observation by month for each cardholder. Next, we're going to summarize that to get a sum, and then we're going to filter to get only the violating one's, to get only those that are above $10,000. Finally, we'll summarize one more time, and here we're going to count the number of violating transactions, and then we're going to get the total spent, which is just the sum of the amount. Additionally, our third one, as we did in Alteryx, is to get the average per month, and that's just the average rather of the sum of the amount. Then we'll arrange descending by count of months.

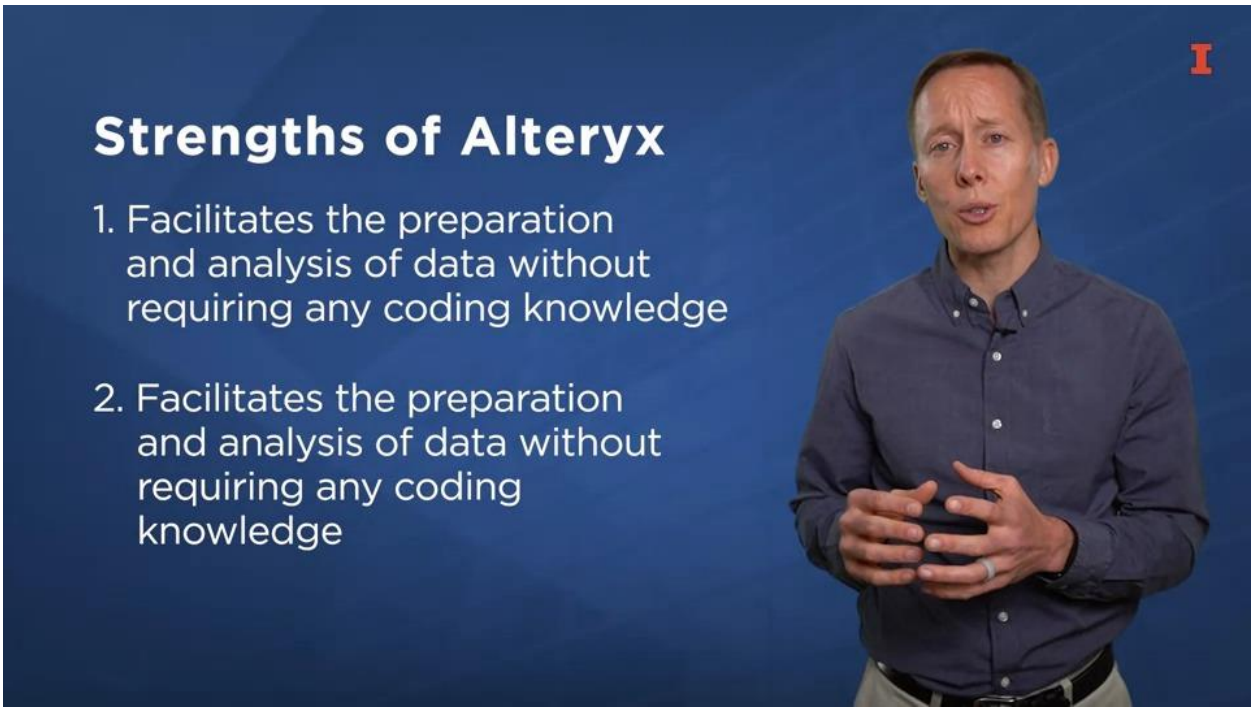


Again, this is the same analysis we saw in Alteryx, where we have Duckworth with 11, so every month but one, and a total spent in those violating transactions of $694,000, and an average of $63,000. Edwards is big, and so on. That's our RStudio. We've gone through and replicated everything we did in Alteryx.

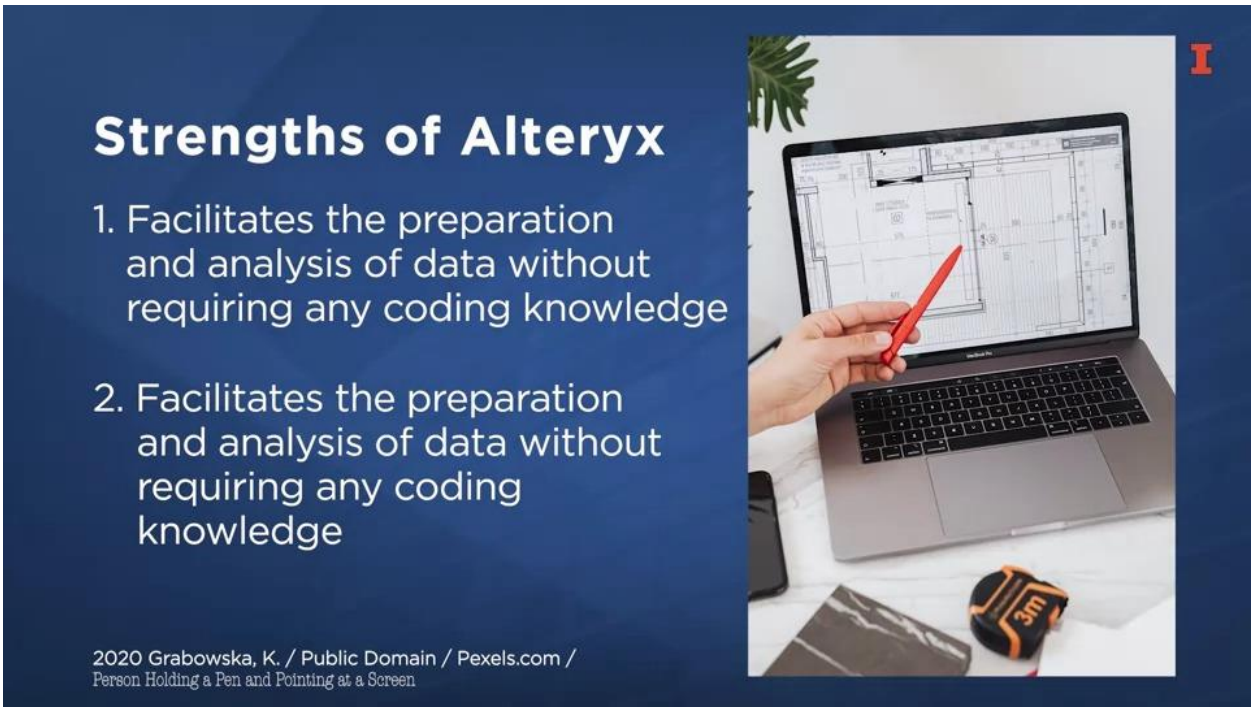
In this video, we cleaned and explored data from real P card transactions from Oklahoma State University and tested the three key controls that they have for controlling spending with purchase cards. This video replicated analysis that we did in Alteryx. My hope is that doing these analyses in both tools will help you improve your skills and let you see which of these tools is most effective for you.

Lesson 2-14: Introduction to Alteryx

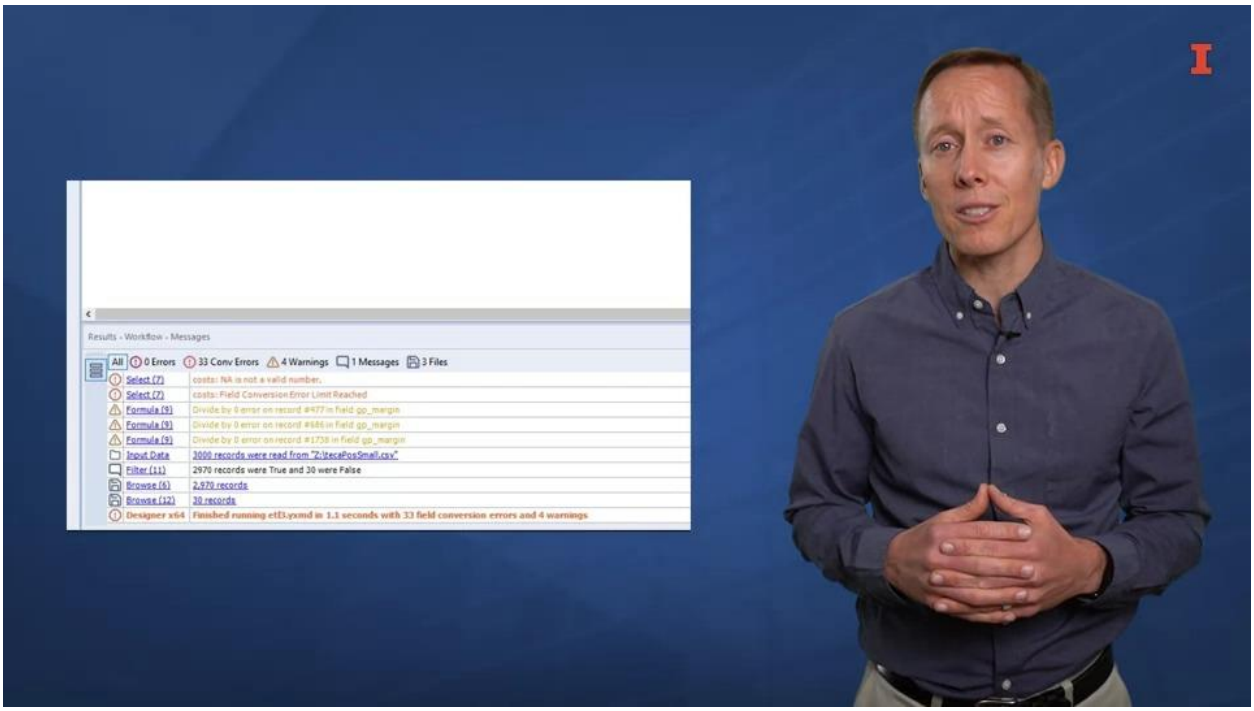
[Music]



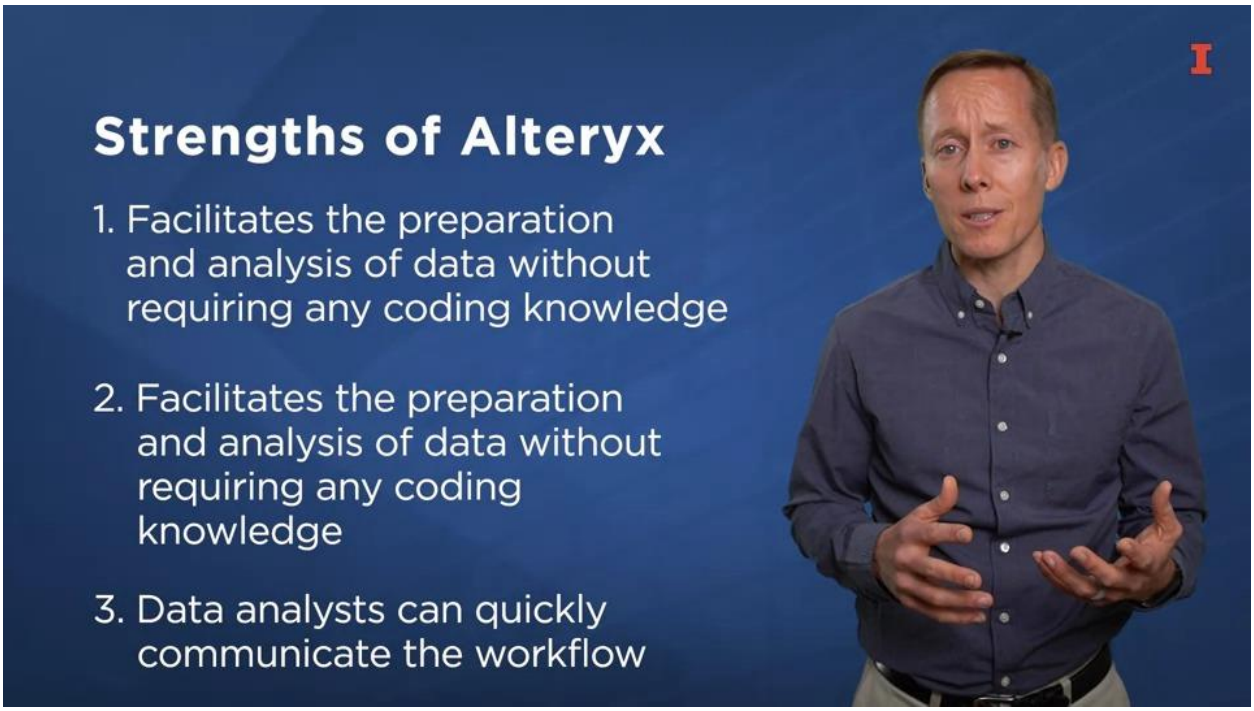
In this lesson, we will introduce you to Alteryx by giving you an overview of what it does and what makes it stand out from other data analytic tools like Power BI, and R. First, Alteryx facilitates the preparation and analysis of data without requiring any coding knowledge. If you can envision the steps that need to take place in the workflow, then you can turn that vision into a reality by dragging pictures that represent the workflow into the right place, connecting some arrows and then configuring the specific details of that workflow. Second, because Alteryx allows users to create a visual representation of the data analytic workflow, data analysts can quickly communicate the workflow with others who can then replicate it quite easily. You could say that Alteryx is to analytics as blueprints and CAD software are to construction.



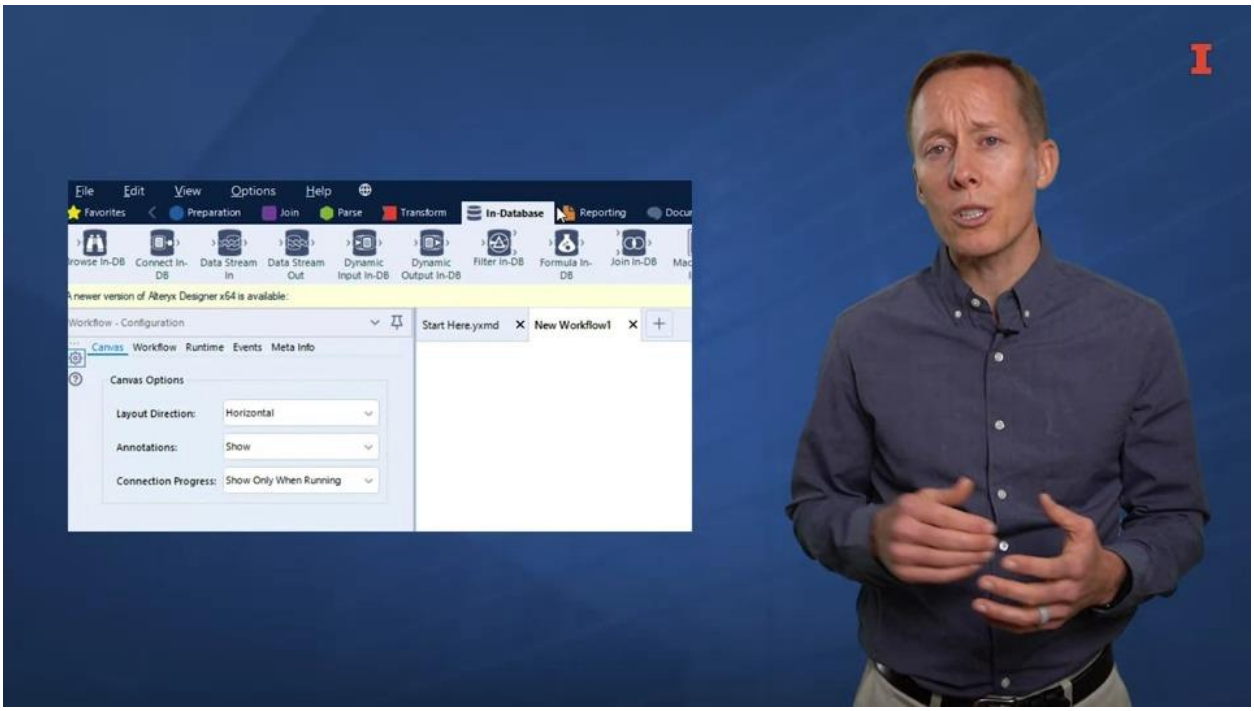
Specifically when communicating instructions for a new building, the architects may create a blueprint or image of how they envision the building to be. That image can then be used to communicate to the builders any details, such as the height and width of the building, as well as the angles and curves, the location of the doors and windows and so on. The builders can then use that image to aid in the creation of the finished product, the actual building.



If we use that process as a metaphor for data analytics, Alteryx allows a data analyst to represent the data analytic workflow with a type of blueprint. This data analytic blueprint can then be used to quickly communicate to others the steps that need to be taken to prepare the data, such as whether new columns need to be created, specific rows or columns need to be selected and datasets need to be joined. With the click of a run button, Alteryx can then convert that blueprint into the finished product, which could be a new dataset, a table, a chart or a report. In addition, that blueprint can then be used by other data analysts to replicate or extend the workflow.

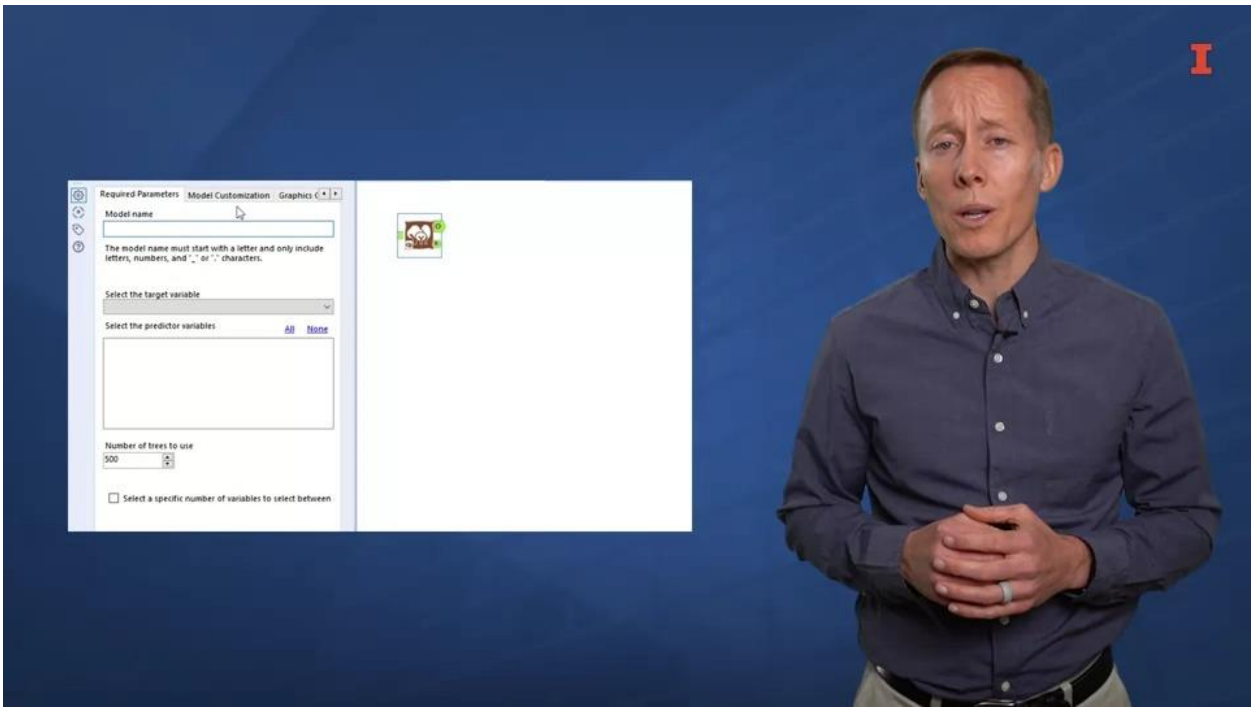


Now, even though Alteryx is great at visualization, data analysts still have to understand the data and data analytic workflow. Alteryx just makes it easier to do so. Nevertheless, Alteryx can help you learn about the data analytic workflow, which is a third strength, the visual representation of a workflow instructs those who are new to the discipline about the steps in the data analytic workflow.



Moreover, the layout of the tools in the tool palette describe the steps that need to be taken for data wrangling and preprocessing and the order in which those steps often occur.

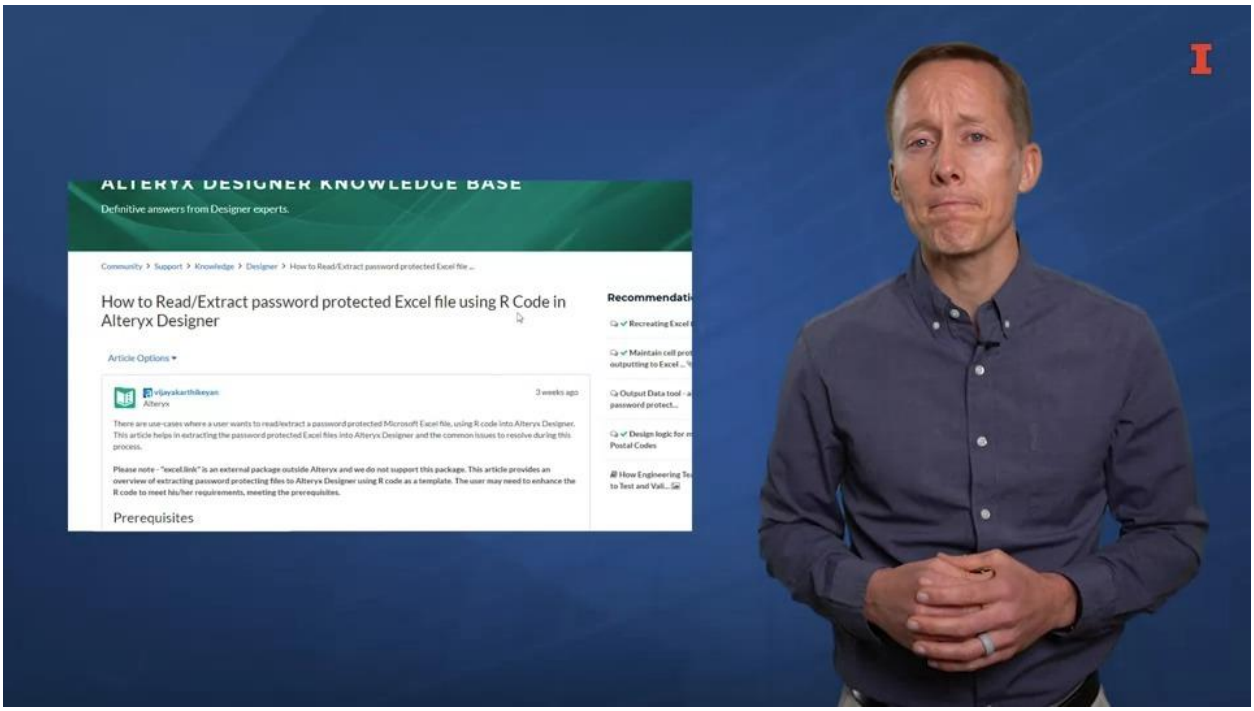
Alteryx is not just for data assembly and exploration, but it is also notable for its ability to facilitate the process of creating a workflow for data modeling with advanced analytic algorithms. As you learn more about that process, you should be able to apply that knowledge very easily with Alteryx.



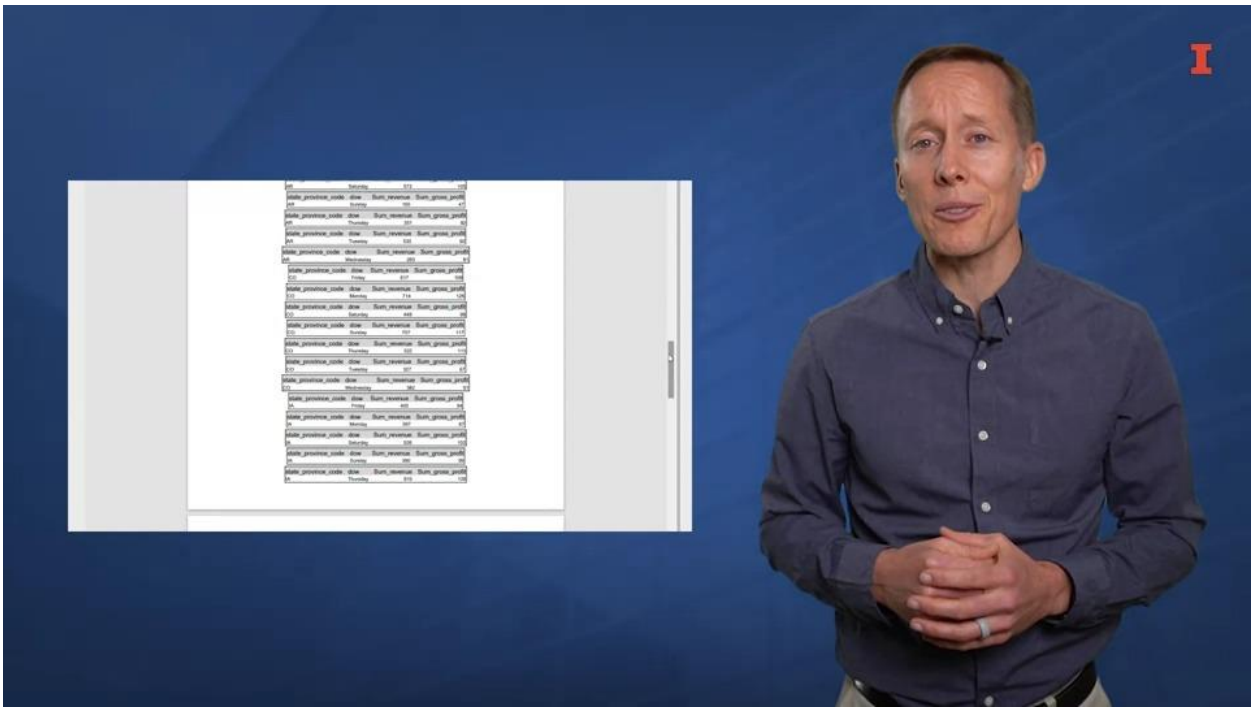
As with data wrangling and preprocessing, I think Alteryx can even help you learn about the data modeling process by exploring the predictive AB testing, predictive grouping, and prescriptive tools in the tool palette. For instance, you can explore the different types of predictive models as well as the parameters for how those models work.

The relative strengths of Alteryx are that it allows data analysts to perform data analytic tasks without coding, visually communicating the data analytic workflow, and serving as an instructional tool for learning about the data analytic workflow.

How does Alteryx do in terms of ease of use and built-in help? In short, Alteryx holds up really well on both of those dimensions. In terms of ease of use, it's a whole different ecosystem from Microsoft Office products. As such, it's initially less intuitive than Power BI for those users who are familiar with Microsoft's products. However, it doesn't take long to understand the Alteryx way of thinking about a data analytic workflow. Once you figure that out, you can easily explore all of the tools that it offers and apply it to your own projects.



One of the reasons why it's easy to develop an Alteryx manner of approaching problems, is that Alteryx has a good amount of built-in help, as well as a strong community of users. The help tools are very accessible, which makes it easy to quickly search for how to complete a process. Furthermore, there are many built-in examples that you can explore, which allows you to learn about processes that you may not even have known that you wanted to complete. Finally, because there's a strong community of users, you can probably find answers to many of your questions outside of the Alteryx tool with an Internet search engine.

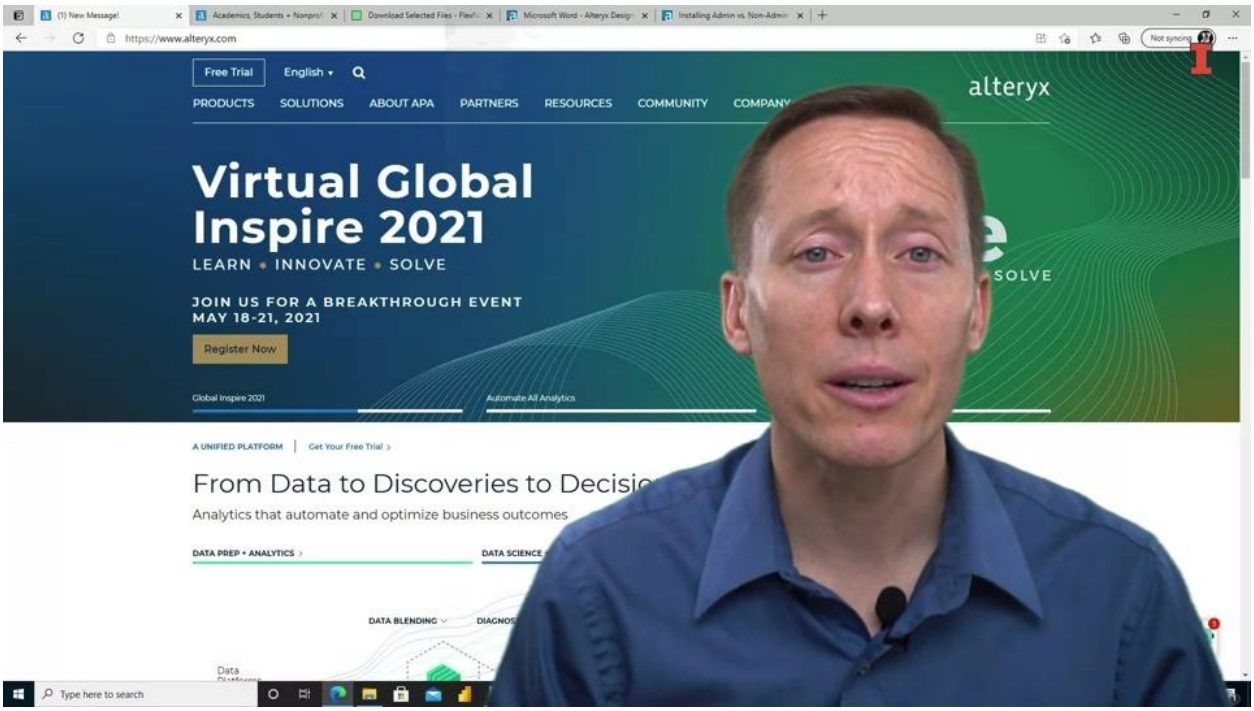


Alteryx is not as strong in visual exploration of data. You can use Alteryx for exploratory data analysis. But that seems to be more of an afterthought rather than a primary focus. You can create a limited number of charts as part of the workflow, but they don't look great and are not front and center. I also find that it's pretty clunky for interactive visualizations and dashboard creation. Alteryx uses Plotly, an open-source third-party provider for interactive charts. Plotly is great, but the way you use it is inconsistent with the way you use the rest of Alteryx. The plots don't always render within the area provided and are not easy to access as they are with say, Power BI. Creating reports is also not a core strength of Alteryx. The good news is that you can create reports and plots, but it may be more effective to send the data into another tool like Power BI, or are for the plotting and reporting.

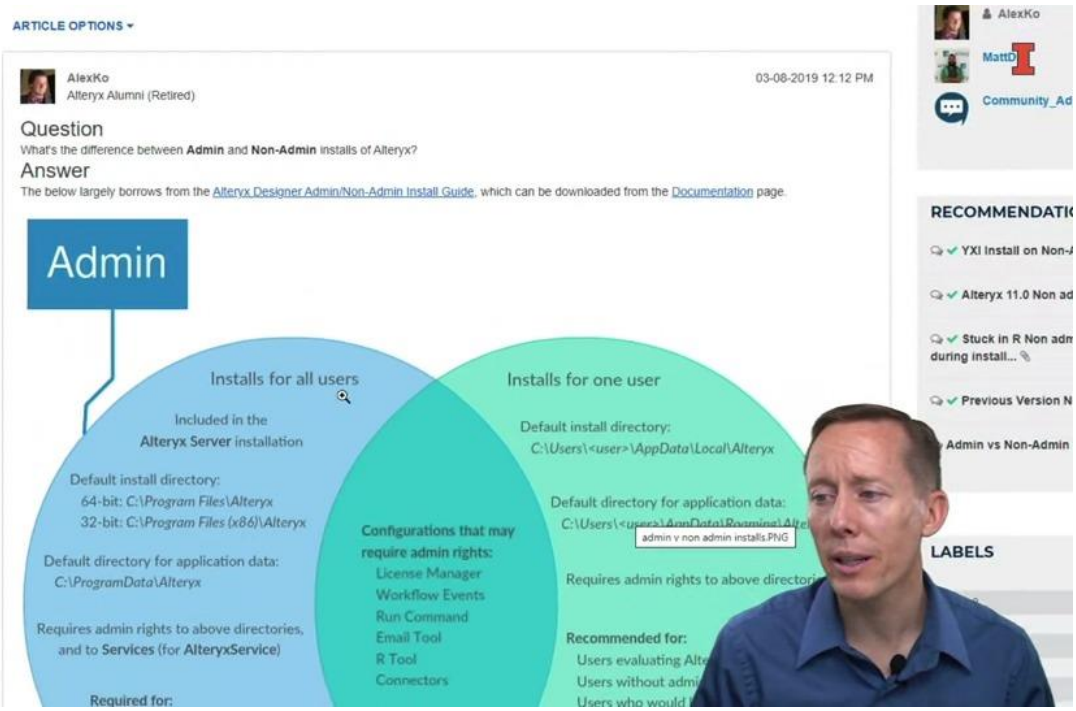
Another weakness is that Alteryx only works on Windows operating systems. For those using a Mac or Linux operating system, you have to go through a few extra steps, such as partitioning your hard drive or installing some other software like virtual box. In short, despite a few limitations, I can see why Alteryx is a popular data analytic tool. It is really strong at creating a workflow that can be easily understood, shared with others, replicated, extended and automated. While it's not great for data visualizations and reporting, it could easily be paired with another tool like Power BI or R to round out the rest of the steps in the data analytic workflow.

Lesson 2-15: Installing and Getting Started with Alteryx

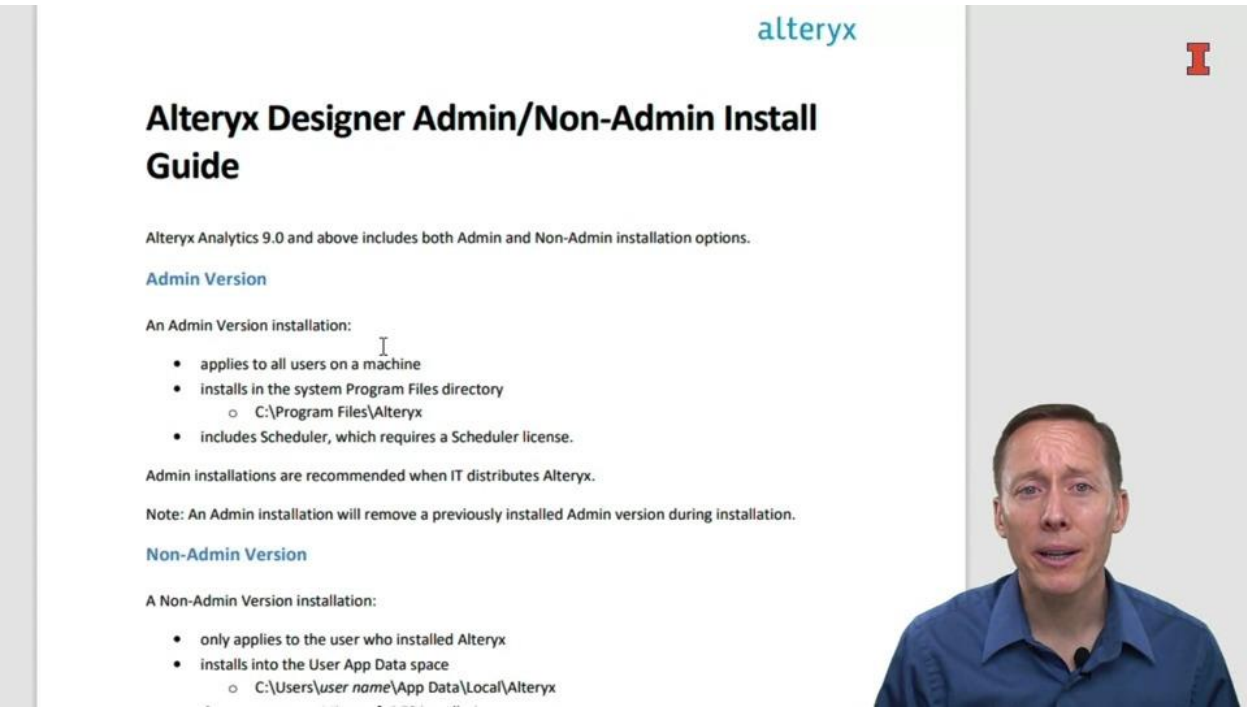
[Music]



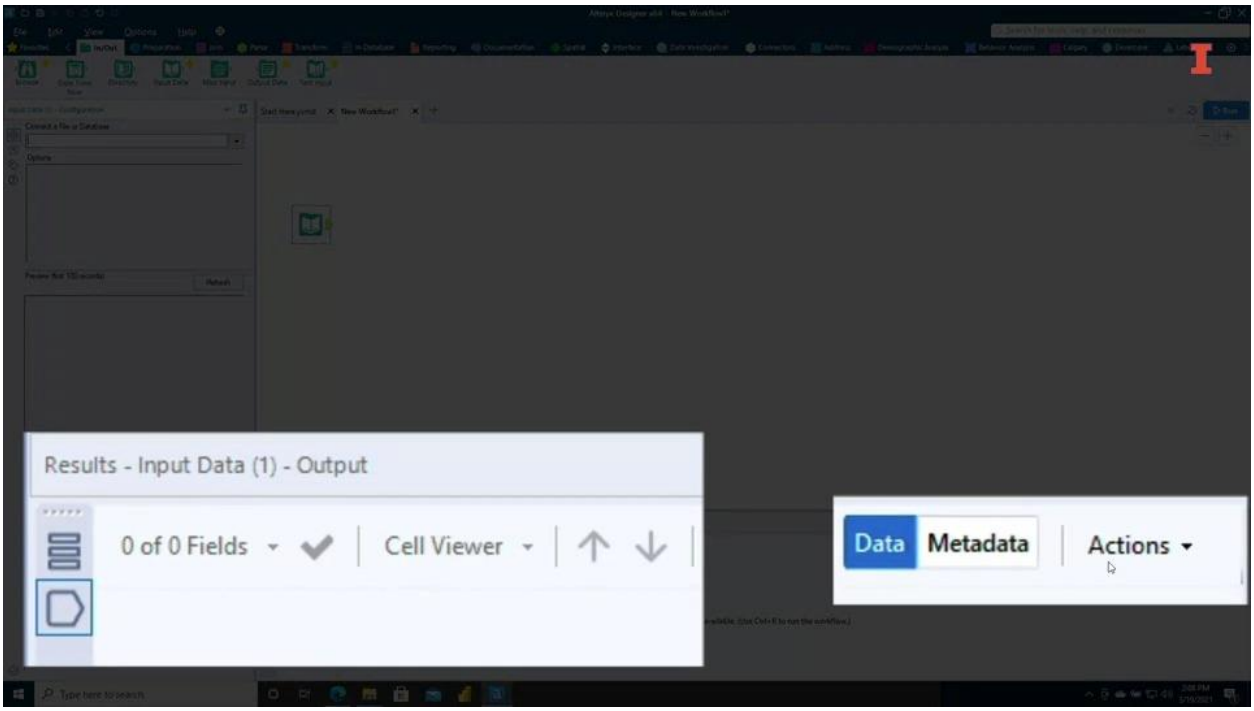
In this lesson, we will talk about how to get started using Alteryx. By the way, the product's name is pretty clever. First, you can think of the way it's spelled as the word alter and the letters y and x. Those letters, y and x, are common letters used to represent dependent and independent variables in supervised machine learning. Second, the pronunciation that I favor, Alteryx, is a reminder that it can do some pretty nifty tricks or maybe even all tricks. To start using Alteryx, the first thing you need to do is make sure that you have a Windows operating system. If you're using a Mac, like me or some flavor of Linux, then you'll have to find a way to use a Windows operating system on your machine. Now I'm currently using virtual box, which is free software that allows me to run Windows 10 and other operating systems on my Mac concurrently with my Mac's operating system. Well, once you have access to a Windows computer, you can install Alteryx by navigating to www.alteryx.com.



Looking at the various products they have on this website, you can see that they have a variety of different products here. The one that we're going to focus on is the designer product. Also notice that you can get a free trial. There are also educational licenses for teachers and students which you can learn more about if you navigate to www.alteryx.com/forgood. You can scroll down and see that there are links for students, for educators, for non-profits, for K-12, and so on. You'll have to decide which link to click on and go through their verification process. Once you get to the point where you can download a product, you'll probably arrive at a web page that look something like this. If you want to follow along exactly with me, then select Alteryx Designer, admin version. The non-admin version is only for a specific user of the machine. For our purposes, I don't think that you'll notice much difference between the admin and non-admin version. I don't think it's a huge deal. Since we are going to focus on using Alteryx for extraction, transformation and loading, or ETL and exploratory data analysis, or EDA, we won't worry about the predictive tools that you see here. You can find out more about the different products such as admin versus non-admin and the designer products if you go to a community website for Alteryx. There is a strong community of Alteryx users. For instance, here you can see a nice Venn diagram of the difference between admin and non-admin users.

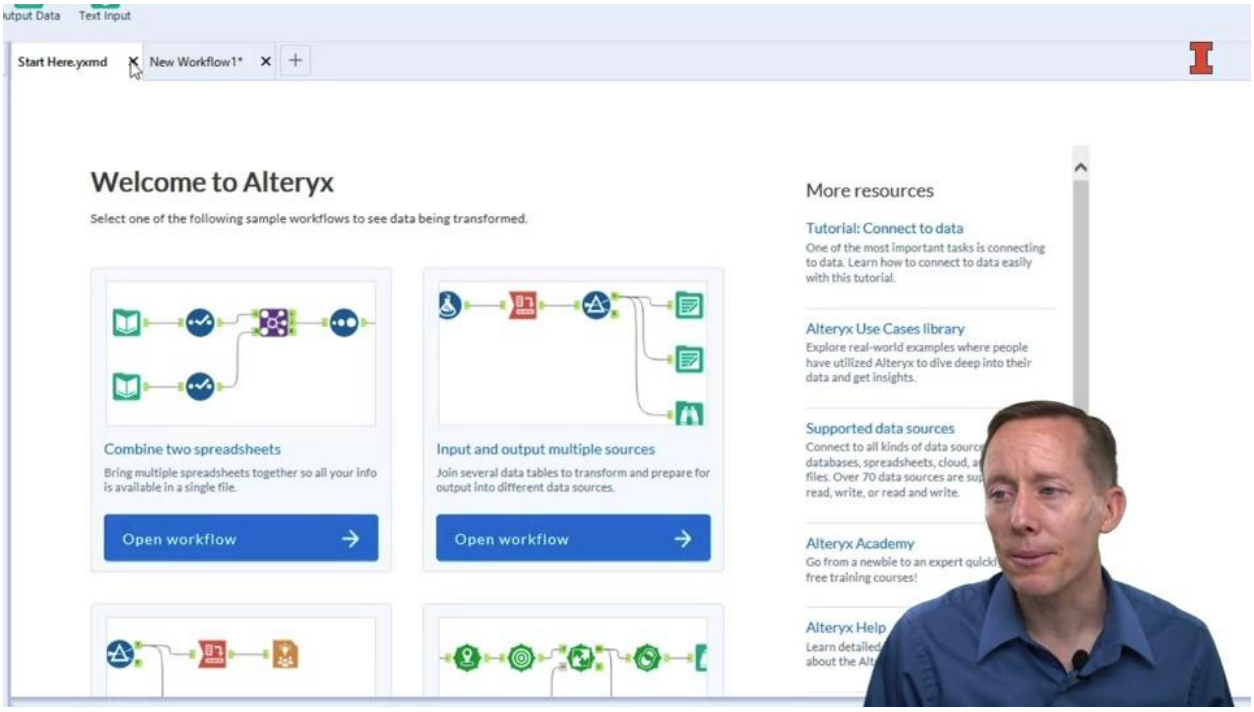


There's also a link to an install guide that also summarizes the difference between the admin and non-admin versions. Now, I suspect that to get the full functionality, that would make Alteryx a great tool for a team of people, you would need to pay extra to get licenses. Anyway, once you've installed Alteryx, let's open it up and explore the layout.

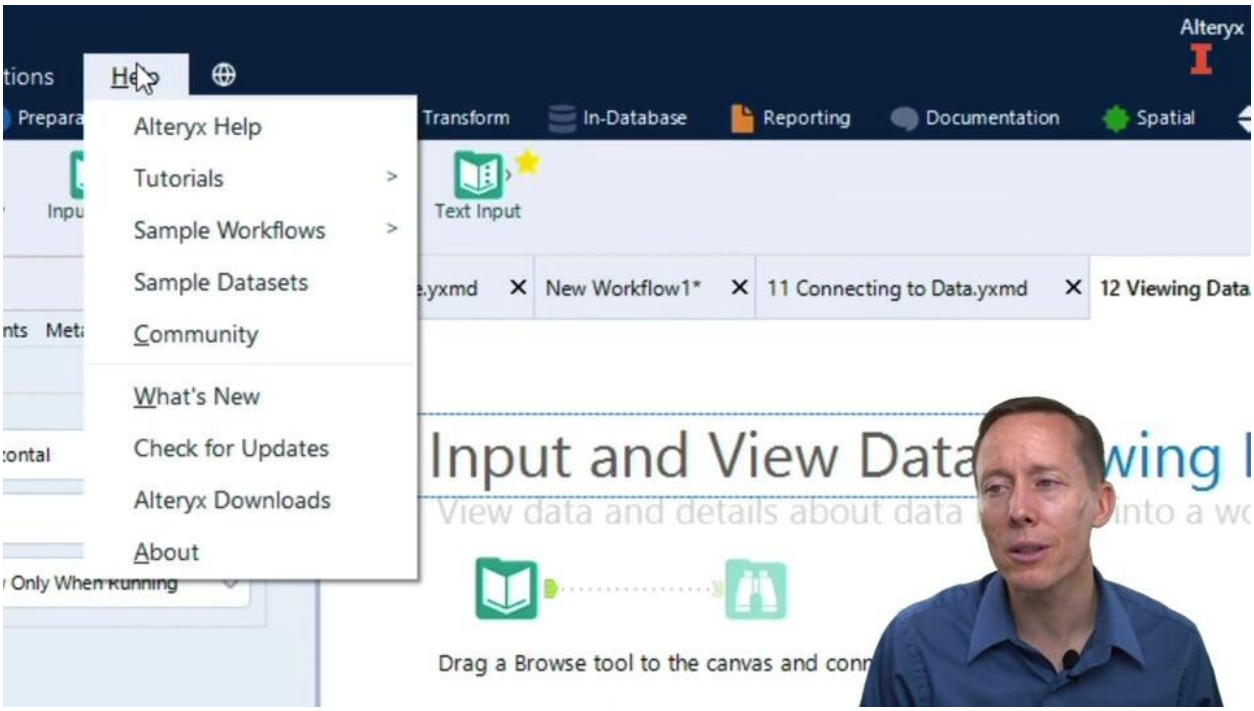


I've already opened it up here, and it'll look just like this and the layout is very different from a typical maybe Microsoft product environment. That's a good thing because it's important to remember that Alteryx is doing something much different from Power BI or other Microsoft products. It's focusing on visualizing the data analytic workflow and not necessarily on visualizing the data itself. Let's talk about the main elements of this Alteryx environment. Along the top here you've got some menu items which are pretty common for a lot of different products. We've got a File menu item where we can create a new file, open a recent file or workflow as they're called save or Save As and then print. We've got Edit to Undo or Redo to Copy Paste. We've got View to customize what this environment looks like. We've got Options which allow us to do a variety of things including run a workflow and importantly, there's also a shortcut to run or some shortcut keys like Control R. That's important because running a workflow is something that we'll do a lot. Then we've got to Help tab, which is really useful and we'll come back to that. Then finally, you can select if you want to view Alteryx in one of about eight or nine different languages. The next thing that you see is what's called the tool palettes, like another menu along the top here. These main elements of the tool palette up here are different major processes of a data analytic workflow. This first one here, Favorites is where you can pin some common processes that you will use a lot and you can customize this. You can see that these processes are color-coded in a way that corresponds to the major process that they relate to. These green ones are for reading data in or saving data out. Blue ones are for data preparation, purple for join and so on. It's actually pretty interesting if you want to just learn a little bit about the data analytic workflow process, you can just explore these different processes along this tool palette here and see actually a common order in which these processes will occur as well as

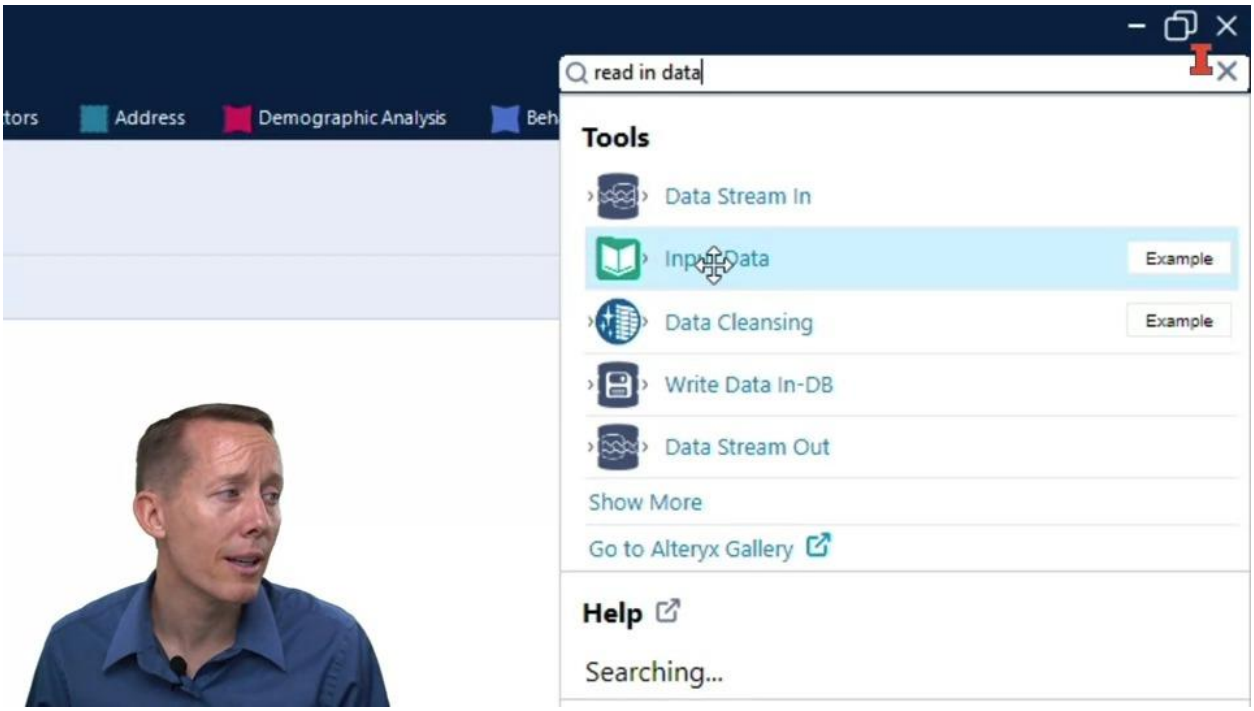
some sub-processes in there. Anyway, there's the tool palette and essentially what you're going to do is oftentimes the first thing is that you will need to input data, connect to a data source. The idea is that this main area here is the Canvas and you will drag elements from the tool palette onto the Canvas to create your data analytic workflow. Then you can configure these elements by specifying different details like, in this case, how you want to connect to a data file in this configuration pane on the left sidebar. Then below is where you can see a preview of the data as well as the results of the workflow that you have run and a few other things like the metadata. Now, actually when you first get started,



there's the Start Here.yxmd, and a.yxmd file is an Alteryx workflow file. This is actually very helpful because you can open sample workflows that correspond to some common data analytic tasks like combining two spreadsheets or inputting and outputting multiple sources. That can be very helpful.



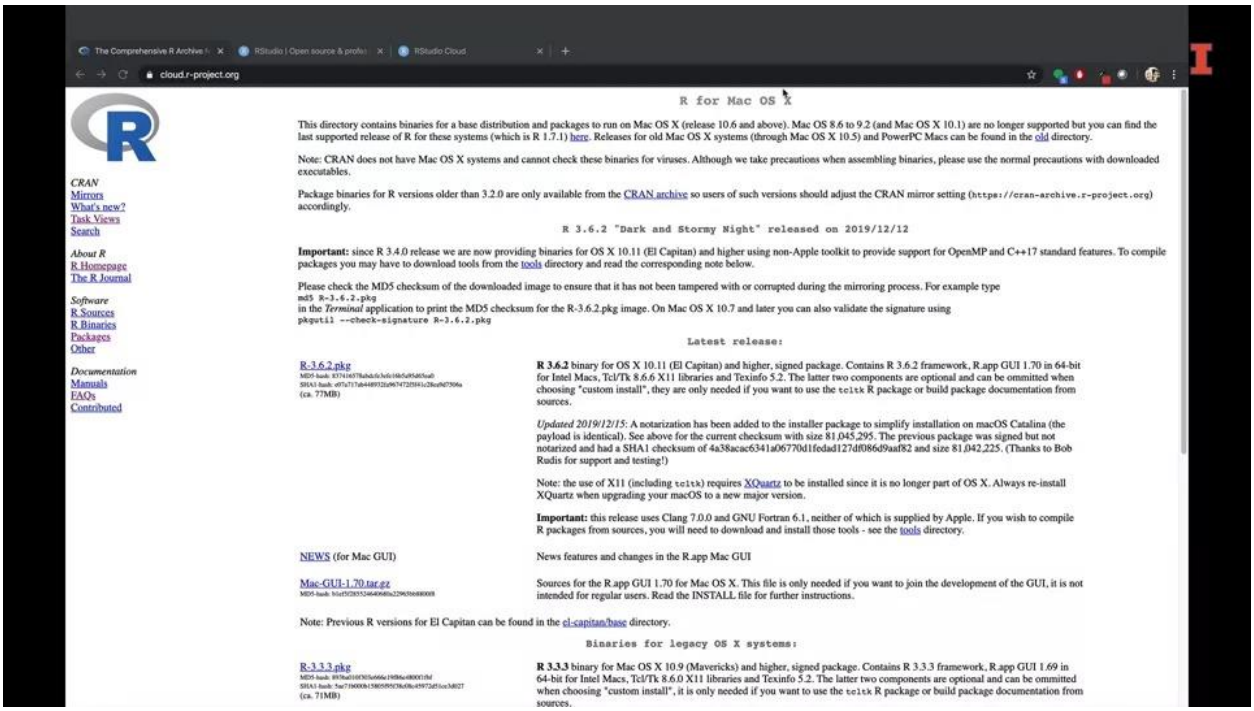
Now, I want to go back up to this Help menu item up here. The Help is so important I think because it allows you to take a little bit of what we'll talk about in these videos and then from there you can start exploring and learning about a lot of other different things. I want to show you just two different help items here. The first one is this tutorials link. If you click on tutorials, you'll see some major elements of a data analytic workflow and if you start at the top here, this will get you started right from the very beginning. You can connect to data. What it will do is it will actually open up an Alteryx workflow file and it shows you how to do that. Essentially, as I just showed you, you drag input data down into here and it gives you a description of what that does, some additional resources, and then you can go to the next step, which might be view data. You click on that and it opens up a new Alteryx workflow file and it shows you what that would look like. You start by reading in data, then connecting it to browsing the data. There's one very useful Help item there is to use these tutorials. There are others in there,



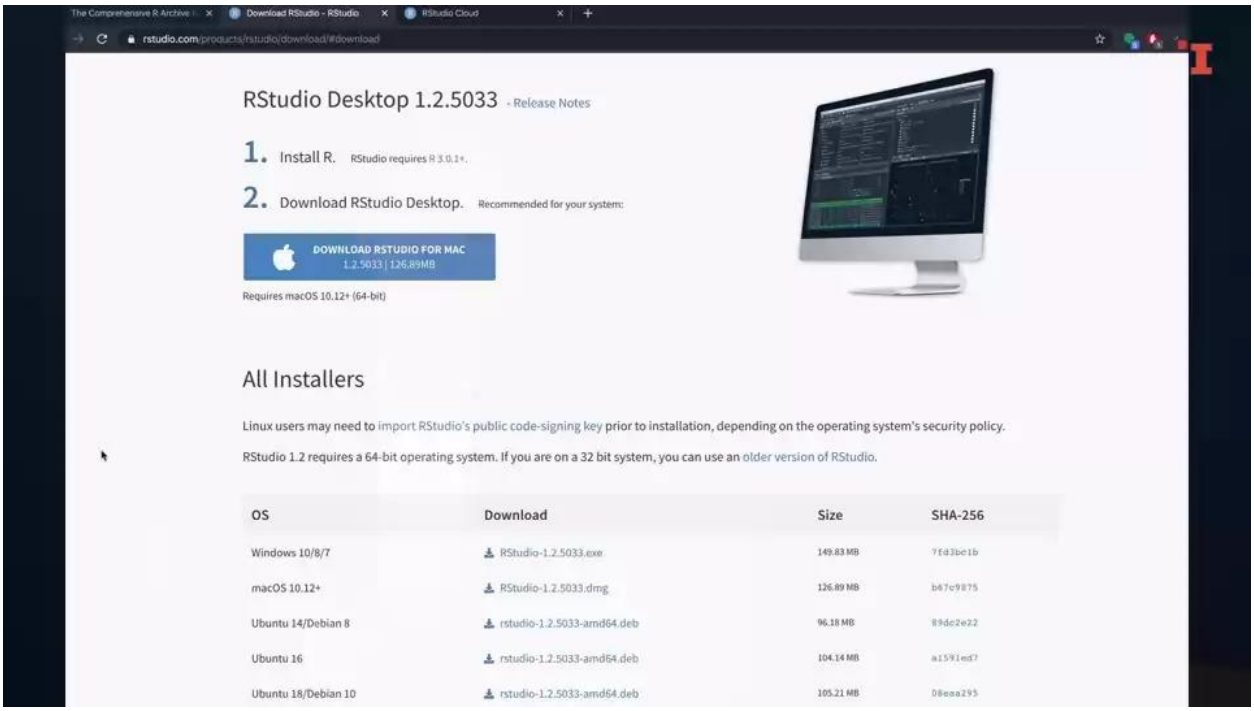
but the other one I want to show you is just the search toolbar at the top. If you just want to type for something that you're interested in doing, like read in data, then you can see links for a variety of things, including Tool processes from the tool palette that will help you do that and links to examples. There's an introduction to Alteryx.

Lesson 2-16: Accessing and Installing R and RStudio

[Music]

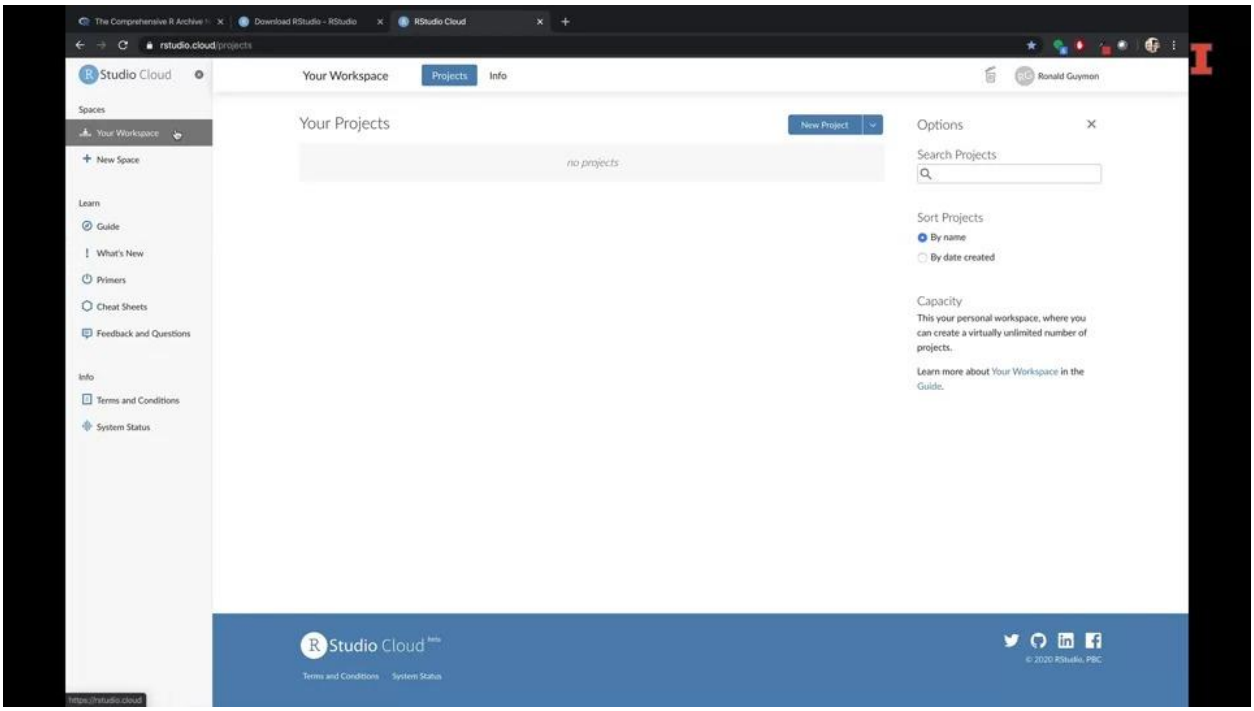


In this video I'll show you how to start using R on your own machine. There are two ways to do this, the first way is to install R and RStudio on your own machine. So let's walk through how to do that. First of all, navigate to r-project.org and browser in the download subsection on the left hand side select the CRAN link. From there you can select a server that's closest to you. But the easiest thing is to select the link under O-Cloud which will automatically redirect you to a server that's closest to you. Then in this top box you'll want to identify the version of R that matches your operating system. So if I'm on a Mac I'll select the link here. At this point you can decide which version of our or you want to download. You can download some older versions but I recommend that you download the latest version which would be right here. So once you click on that link it will download and installer. Once it finishes downloading the installer you can double click on it and walk through the steps to install. At that point you'll be able to open R in your programs on your machine.

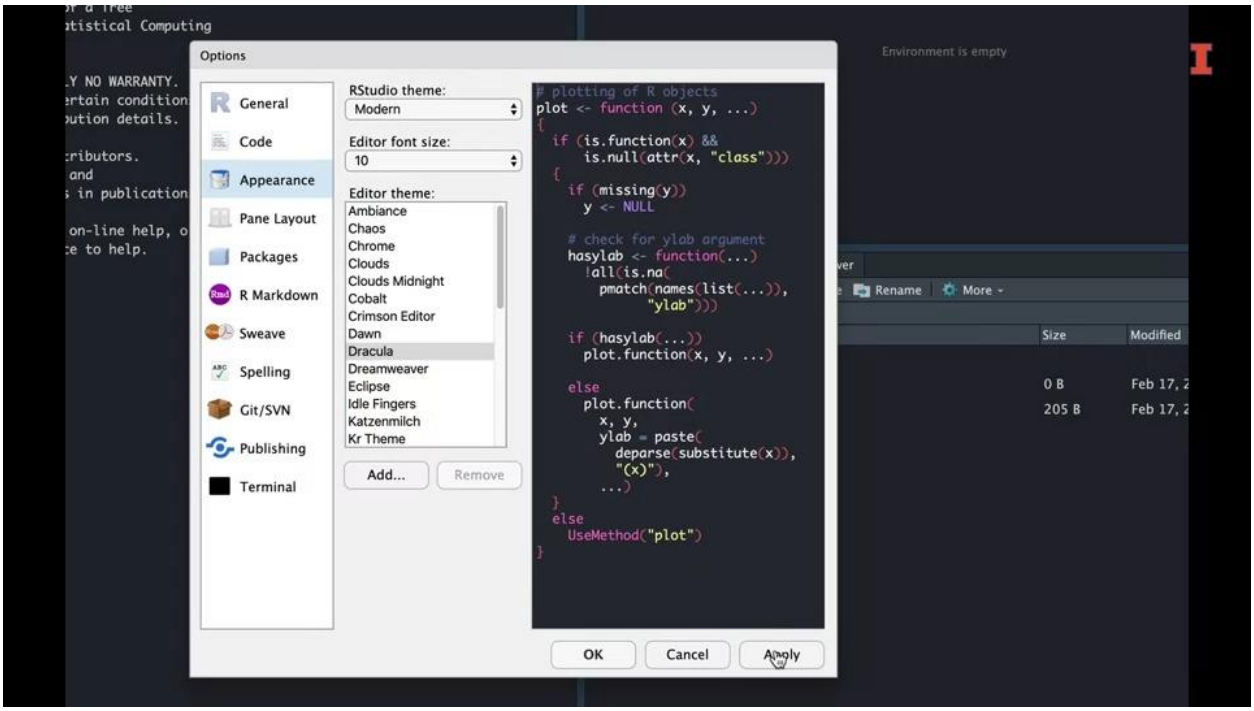


It will open up an IDE or integrated development environment that looks like this. It's pretty basic but it works but I want to show you how you can use RStudio that looks like this. So to install RStudio navigated to rstudio.com and a web browser. In this products menu item at the top go over to the far left and under Open Source, select RStudio the premier IDE for R, select that. Now you have two versions of RStudio. We're going to go with RStudio desktop. Before you click on that take a minute to pay attention to the video tour of RStudios IDE. It's a brief two minute video and it does a great job of giving an overview of what each of the different major elements of the RStudio IDE does. I highly recommend that you watch that. Once you've watched the video go ahead and click on the RStudio Desktop and then it will give you several options. The first choice that you'll have to make is if you want the open source edition or a paid version. So we want the open source edition because it's free. So we'll download RStudio Desktop. It takes you to this page where you can select a free or a paid version of RStudio desktop. And go ahead and go to the free version. It should then detect automatically what type of system you're on and direct you to the version of RStudio that matches your operating system. So at this point I go ahead and click on this button here. Once it finishes downloading the installer I will double click on that and go through the steps to install RStudio.

Now invariably there will be some of you who have a problem in one way or another. Installing either R or RStudio onto your machine, I would highly recommend trying to figure out the problem and trying to resolve it. We won't take time in this video to troubleshoot potential problems, but if they come up, do your best to resolve them by searching online.



Now, if you don't want to resolve this issue or you don't have time, the second option for using RStudio on your machine is to use RStudio Cloud. Now this is a product that's provided by RStudio and you can find it by navigating to rstudio.cloud in a web browser. It's a free resource, but you have to sign up for it. So go ahead, sign up for it and then log in. Okay? Once you've signed up and logged into RStudio cloud, you'll notice that you've got this left hand sidebar over here. And you want to go ahead and just maybe review what's over here briefly. The main thing is we want to create a project. So go ahead and click New Project.

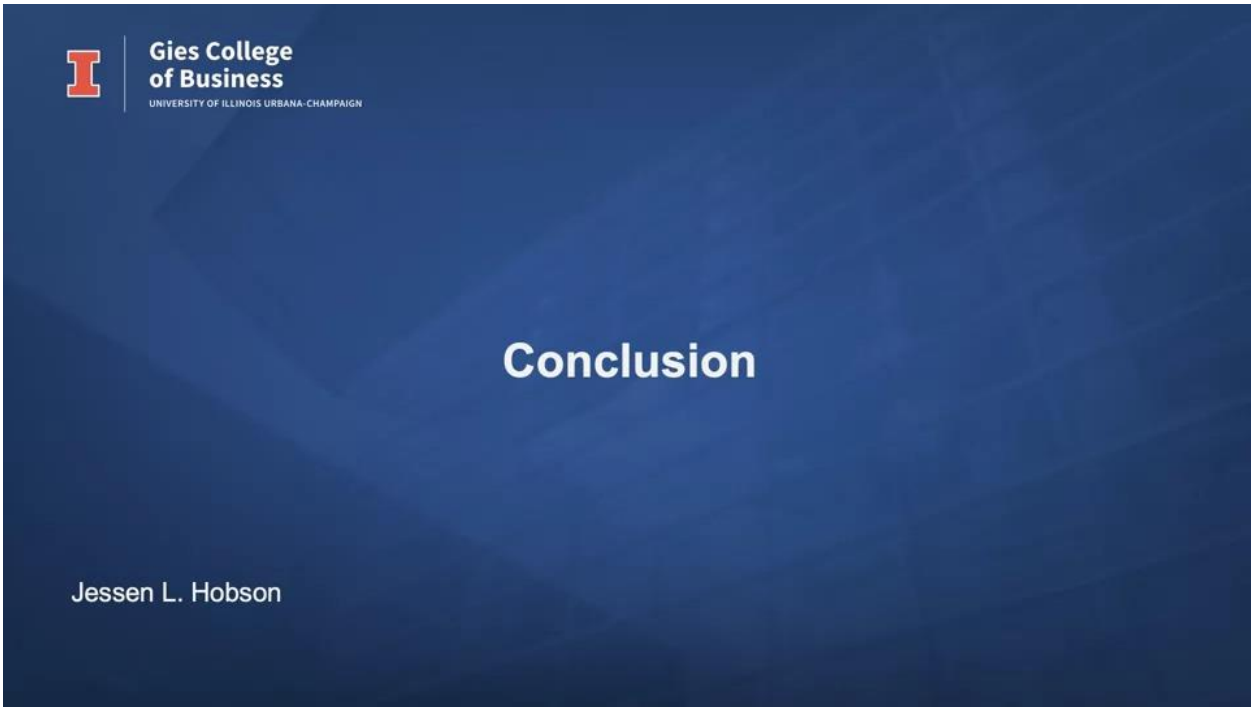


All right, once you've created a project, you'll see that you've got an environment that looks basically identical to the desktop version of R. With the exception that you've got a web browser information along the top and some RStudio Cloud information on this left side bar. So at this point I want to show you how you can adjust some critical settings that will help you be able to use RStudio effectively. So go up to the Tools menu item, whether you're on a desktop or RStudio cloud and navigate down to Global Options. You can see there are a bunch of options along the side. So at some point you'll want to explore those at this point, select appearance. Let's focus on these two boxes, the Editor Font Size. So if your screen is really far back or based on your vision, you can select a font size that's bigger or smaller than what the default setting of 10 is. You can also select a theme. Instead of having the white background, for instance, you can select dracula and it has a black background and light to color the text. Then you can select Apply and OK? And you can see that it changes the way RStudio is visualized.

So those are some important settings that I think you should know about so that you can effectively use R and R studio on your own machine

Module 2 Conclusion

Controls and Auditing: Conclusion



[Music]

In this module, we explored the issues of technology and analytics within the audit function.

Auditors are accountants that are called upon to review the financial numbers of organizations, as well as their internal controls. Thus, auditors play a vital role in making sure companies follow the rules, do not commit fraud, do not inject misstatements into their financial reports, and don't mislead users of the accounting statements. This helps ensure that the capital markets are safer for the world. Technology plays a vital role in auditing. Auditors are called upon to examine big data and to explore that big data in important ways. Auditors must leverage analytics if they want to be successful at finding the mistakes of others, and revealing intentional and unintentional errors. This module helps you to understand the environment of the audit and the importance of internal controls. It then demonstrated one way that analytics can facilitate the testing of internal controls through a realistic case using purchase card transactions at Oklahoma State University. Using both Alteryx and RStudio, the solutions to the case demonstrated the importance of analytic skills. This case provided you with one concrete application of analytics in auditing. There are many interesting and useful applications of analytics in this area, and I encourage you to explore those and to learn more, using what we've talked about in this module as a framework and a baseline to build upon. The skills learned here are applicable not only to auditing and company controls, but in any area of analytics. Thus, the skills learned here will help you in any area of business.