Data Analytics Readings

Session 1: Validity and Causality

Types of Validity

- Statistical Conclusion Validity: This refers to the accuracy of inferences about the relationship (covariation) between a treatment and its outcome. It focuses on whether the variables are statistically correlated.
- Internal Validity: This addresses whether the observed covariation between the treatment (A) and the outcome (B) is a result of a causal relationship between them. It's about establishing that the treatment, and not some other factor, caused the observed effect.
- Construct Validity: This is the validity of inferences that generalize from the specific elements of a study—such as the people, settings, and treatments—to the broader theoretical constructs those elements are meant to represent.
- External Validity: This type of validity concerns the generalizability of a cause-and-effect relationship. It asks whether the findings of a study hold true for different people, in different settings, and with different variations of the treatment and measurement.

Causality Relationships

The relationships between constructs (the unobservable, "what counts") are considered causal, whereas the relationships between their proxy measures (the observable, "what gets counted") are merely statistical. This highlights that a statistical association does not necessarily prove a causal relationship because other factors, or "confounds," may be responsible for the observed correlation.

1. Temporal Precedence: The cause must precede the effect in time.
2. Covariation: There must be a relationship between the cause and the effect, meaning that a variation in the cause is related to a variation in the effect.
3. No Plausible Alternatives: There should be no other plausible explanation for the effect other than the presumed cause.

Libby Diagram

The purpose of Libby Boxes, or diagrams, is to distinguish between what "counts"—the underlying, unobservable constructs—and what "gets counted"—the observable proxy measures. The framework, rooted in the predictive validity framework, helps explain that a causal relationship exists between constructs, while only a statistical relationship exists between their measures. The diagrams are considered a useful tool for understanding managerial reporting.

The Boxes and Arrows

- Top Boxes (What Counts): The top boxes represent the underlying, theoretical constructs that have a causal relationship.
- Bottom Boxes (What Gets Counted): The bottom boxes show the proxy measures used to represent the constructs.
- Dashed Lines: The vertical dashed lines connecting each construct to its proxy measure represent measurement error. This highlights the fact that the measure is not a perfect representation of the underlying construct.
- Omitted Variables: A fifth oval box can be added to the diagram to represent omitted variables that influence the proxy measure for the effect. These variables, such as "intelligence," "prior exposure," or "stress," can affect the measured outcome without being part of the hypothesized causal relationship.

Session 2: What is Statistics

Statistics Help Us Process Data

The author highlights that statistics can summarize a large amount of information into a more manageable form, allowing us to make sense of complex datasets.

- Descriptive Statistics: These methods, like finding the mean or median, provide a way to describe a dataset concisely. For example, instead of listing the height of every person in a class, you can simply say the average height is 5'9".
- Inference: Statistics allows us to use a small sample of data to make educated guesses about a much larger population. This is how pollsters can predict election results by surveying a small group of voters.
- Revealing Hidden Patterns: Statistics can uncover relationships and patterns in data that aren't obvious at first glance. For example, a statistical analysis might show a correlation between ice cream sales and shark attacks, but it would also help us understand that a lurking variable (warm weather) is likely causing both.
- Evaluating Claims and Models: Statistics provides a framework for testing whether a claim is likely to be true or if a model accurately represents reality. This is crucial for distinguishing between valid conclusions and random chance.

What's the Point of Learning Stats

- Statistics can be used to understand information that ranges from the trivial, like sports statistics, to the profoundly significant, like the Gini index for income inequality
  - To summarize huge quantities of data
  - To make better decisions
  - To answer important social questions
  - To recognize patterns that can refine how we do everything from selling diapers to catching criminals
  - To catch cheaters and prosecute criminals
  - To evaluate the effectiveness of policies, programs, drugs, medical procedures, and other innovations
- The book also promises to explain how statistics can be applied to real-world situations
- The author's goal is to demonstrate that statistics can be interesting and accessible by focusing on intuition and using real-world examples, rather than getting bogged down in complex mathematical formulas
- Wheelan's core argument is that while it is easy to lie with statistics, it is hard to tell the truth without them

Session 3: Descriptive Statistics

Descriptive Statistics

What are descriptive statistics, generally speaking?

Descriptive statistics are numbers and calculations used to summarize raw data. Instead of looking at a massive, unwieldy spreadsheet of information, descriptive statistics distill that complex data into a handful of simple and meaningful numbers.

Why are they helpful?

Descriptive statistics help us process information and gain clarity from large datasets. They serve to simplify and summarize data, making it manageable and understandable. This allows for easy comparisons, such as assessing the performance of two quarterbacks on a given day or comparing the economic inequality of different countries. They provide a "synopsis of what happened" and frame important issues, even if they don't provide a single "right" answer.

Dispersion

How is dispersion measured?

Dispersion, which refers to how spread out a set of data is, is measured by the standard deviation. This statistic assigns a single number to the dispersion of a distribution around its mean. A larger standard deviation indicates that the data points are more spread out, while a smaller one means they are more tightly clustered around the mean. This measure is what allows us to define a "normal" range for phenomena, like a person's height or a blood chemical count.

What is the normal distribution and why is it useful?

The normal distribution is a common, bell-shaped distribution of data that is symmetrical around its mean. It's used to describe many natural phenomena.

It is incredibly useful because its properties allow for powerful statistical conclusions:

- Predictable Proportions: For any normal distribution, a specific proportion of the observations fall within a certain number of standard deviations from the mean. For instance, roughly 68% of the data lies within one standard deviation of the mean, and about 95% lies within two standard deviations. This predictability is the foundation for much of statistical inference.
- Foundation for Inference: The normal distribution is the cornerstone of the Central Limit Theorem, which states that the means of large, random samples will be distributed normally around the population mean, regardless of the original population's distribution. This allows researchers to use data from a small sample to make reliable inferences about a much larger population.

Assessing School Quality

The core of this example lies in distinguishing the construct from the measure.

- The Construct ("What Counts"): The underlying concept of school quality is the construct. It's a complex, theoretical idea that is difficult to observe and measure directly. A "high-quality school" is an abstract idea that includes many factors beyond simple academic performance.
- The Measure ("What Gets Counted"): In reality, we rely on proxy measures to stand in for this abstract construct. In the case of schools, these measures are tangible statistics like test scores and graduation rates.
- Challenges to Validity: Wheelan argues that relying solely on test scores presents a "dangerously inaccurate picture" of a school's quality. This is because the measure (test scores) is not a perfect representation of the construct (school quality) due to significant outside influences. The example of selective-enrollment schools that admit students based on high test scores further complicates this, as it demonstrates that high scores are an input to the "school quality" measure rather than an output of the school's educational value. This compromises the validity of any conclusion that the school's teaching is what produced the high scores.

Wheelan's critique highlights the potential for measurement error, noting that test scores are influenced by factors like students' parental income and background, not solely by the quality of the school itself. This idea is visually represented by the Libby Box framework.

This example directly questions the construct validity of using test scores as a measure of school quality. Construct validity is the truthfulness of the inference that a measure accurately represents the higher-order concept (or construct) it is intended to capture.

Session 4: AI and Learning

AI Impact Learning

- AI can impact learning by enhancing it or being a detriment
- If it's used as a learning assistant, it can help enhance learning
- If AI is used to just complete assignments, then it is a detriment

Cognitive Offloading

- Cognitive offloading is the process of using external tools or resources to reduce the mental effort required for a task. This is a natural human tendency that allows us to manage complex information without overwhelming our brains.
- Scope of Offloading: Older tools were designed to offload a narrow set of applications or discrete skills. For instance, a calculator offloads mental arithmetic, and GPS offloads spatial memory and map-reading. AI, however, offloads a "much more complex set of processes", such as writing, problem-solving, and critical reasoning across multiple domains.
- Active vs. Passive Engagement: Previous technologies augmented human capabilities, often freeing up mental resources for higher-level thinking. For example, a calculator lets you focus on the problem's logic rather than the arithmetic. With generative AI, the tools are so "autonomous in their execution" that they allow a user's mind to "effectively power down". This reduces the need for deep cognitive engagement and the "productive struggle that drives problem-solving".
- Impact on Skill Retention: By offloading entire tasks, AI risks the "quiet erosion of our cognitive capabilities" and "long-term cognitive atrophy". Studies show that students using AI for math problems saw short-term gains but performed worse without the tool later on. This overreliance can dull skills that once kicked in automatically, making tasks like breaking down problems and organizing thoughts feel "impossibly difficult".

AI Enhance vs Weaken Learning

When AI Enhances Learning

AI can be a powerful tool that enhances education when used thoughtfully and strategically.

- Personalized Learning: AI platforms can analyze student data to identify learning gaps and adapt content to an individual's pace and style. This customized approach keeps students engaged and can lead to improved academic performance.
- Immediate and Detailed Feedback: AI can automate grading and provide instant, detailed feedback on assignments. This helps students understand their strengths and weaknesses more quickly and allows teachers to assign more writing tasks and offer timely corrections.
- Administrative Support for Teachers: AI can handle routine administrative tasks, such as grading, scheduling, and generating lesson plans. This frees up teachers to focus on what they do best: mentoring, providing personalized guidance, and building crucial relationships with students.
- Accessibility: AI tools with features like text-to-speech and visual recognition can make educational material more accessible to students with special needs, promoting a more inclusive classroom environment.

When AI Weakens Learning

AI can be detrimental to learning when it encourages a reliance on shortcuts and replaces the cognitive "struggle" necessary for skill development.

- Erosion of Critical Thinking: Heavy reliance on AI for problem-solving can weaken students' ability to think analytically and independently. A study found that students who used AI for math problems initially outperformed their peers but scored lower when tested without the AI's help.
- Dehumanization of Learning: Overdependence on AI tutors can diminish the vital role of human teachers in fostering social interaction and emotional development. This can lead to students bypassing the valuable cognitive struggle of forming hypotheses and drawing conclusions on their own.
- Cognitive Offloading: While humans have always used tools for cognitive offloading, AI takes it to another level. Older technologies offloaded specific, discrete skills, but generative AI can offload entire complex thought processes, such as writing and problem-solving, causing the mind to "power down" and leading to long-term cognitive atrophy.
- Bias and Misinformation: AI systems can inherit biases from their training data, which can lead to unequal learning experiences and the propagation of inaccurate or biased information.

Strategies to Enhance Education with AI

To ensure AI serves as a powerful partner rather than a crutch, students and teachers can adopt the following strategies:

- Draft First, Prompt Second: Students should outline their own thoughts and ideas before turning to AI to elaborate or polish their work. This ensures they engage with the problem-solving process and use the AI to refine, not replace, their own thinking.
- Use AI as a Tutor, Not an Answer Dispenser: AI can be used to guide students through a problem step-by-step, providing hints and encouraging them to explain their reasoning. This cultivates lasting comprehension instead of just offering a quick answer.
- Incorporate "Cognitive Forcing": Teachers can use techniques from medicine and aviation, like "diagnostic timeouts," to encourage students to pause and critically evaluate AI outputs before accepting them. This can involve having students restate the main points in their own words or run a mental checklist to check for bias or missing perspectives.
- Create Authentic and Staged Assignments: Teachers should design assignments that require students to reference in-class content, demonstrate their process through drafts, and reflect on how they used AI. This focuses on the development of skills rather than just the final product.
- Open Communication and Transparency: Teachers and students should have open conversations about AI's strengths and limitations. Setting clear policies on acceptable and unacceptable uses of AI helps to manage expectations and promote responsible use.

Session 6: Probability

Challenging the Objectivity of Probability

- Epistemic vs. Aleatory Uncertainty: The article explains that numerical probability is used for two types of uncertainty: aleatory (about the future, or chance) and epistemic (about what we currently do not know, or ignorance). The subjective nature of probability is highlighted by the coin flip example: once the coin is flipped and covered, the probability of heads is simply a measure of your ignorance, not a feature of an unknown future.
- "All Models are Wrong, But Some are Useful": This aphorism, quoted in the article, underscores the idea that even in science, probabilities are not "true" but a product of subjective, if reasonable, assumptions and judgments. For instance, a P value calculation in a clinical trial relies on numerous assumptions about the statistical model that are technically false, though the strength of the evidence may overcome these modeling flaws.
- Pragmatic Approach: Despite these limitations, the article concludes that it is often useful to act as if probability does exist, taking a pragmatic approach.

Critiquing the Misuse of P-values

- P-values are Not Truth or Chance: The most critical principle is that P-values do not measure the probability that the studied hypothesis is true, nor do they measure the probability that the data were produced by random chance alone.
- P-values Only Indicate Incompatibility: P-values only serve to indicate how incompatible the data are with a specified statistical model (the null hypothesis). A smaller p-value simply indicates greater statistical incompatibility.
- The "Bright-Line" Misuse: The widespread practice of using $p<0.05$ as a mechanical "bright-line" rule for justifying scientific claims is strongly discouraged, as it leads to considerable distortion of the scientific process.
- Statistical vs. Scientific Significance: P-values do not measure the size or importance of an effect. A tiny effect can yield a small p-value if the sample size is large enough, which directly relates to the importance of comparing statistical significance with effect size, a key consideration for avoiding the misinterpretation of results.
- P-hacking (Selective Reporting): The ASA warns against "cherry-picking promising findings," or p-hacking, by conducting multiple analyses and reporting only those that pass a significance threshold. This renders the reported p-values essentially uninterpretable and is a form of statistical deception.

Assuming Events are Independent When They Are Not

- This is a critical mistake where people treat the occurrence of one event as having no influence on the probability of another event, when in reality, the events are related. If events are truly independent, the probability of both happening is the product of their individual probabilities.
- The Danger: This error dramatically underestimates the likelihood of multiple connected events occurring.

Not Understanding When Events Are Independent

- This error, often called the gambler's fallacy, occurs when people believe that the outcome of a past event somehow makes a subsequent, unrelated event "due" to happen.
- The Concept: For truly independent events, the outcome of the first event has no effect on the probability of the second event.

Clusters Happen

- This is the misinterpretation of natural, random variation as evidence of an underlying cause, often driven by selection bias.
- The Concept: An event may be individually improbable, but when considering the vast number of opportunities for that event to occur, it is not improbable that it will happen somewhere. People tend to focus on the area where the anomaly occurred while ignoring the millions of places where it did not.

Reversion (Regression) To The Mean

- This is the phenomenon where any observation that is an extreme outlier (unusually high or low) is likely to be followed by outcomes that are closer to the long-term average (the mean).
- The Concept: Performance often consists of underlying talent plus an element of luck (good or bad). An extremely good performance is usually partly due to good luck, and the next performance is unlikely to replicate that luck, thus moving back toward the normal average.

Session 7: Interpretation of P-Values

Hypothesis Testing (Null vs Alternative)

Hypothesis testing is a formal statistical process used to determine if there is enough evidence in a sample of data to reject a claim about a larger population. This process is structured around two competing statements: the null hypothesis and the alternative hypothesis.

Null Hypothesis (H_0)

- The null hypothesis (H_0) is the initial assumption or starting claim that a researcher attempts to disprove.
- The Status Quo: It generally represents a statement of no effect, no difference, or no relationship.
- The Burden of Proof: In the framework of hypothesis testing, the null hypothesis is assumed to be true until the data provide overwhelming evidence to suggest otherwise. It's analogous to the presumption of innocence in a court of law.

Alternative Hypothesis (H_a or H_1)

- The alternative hypothesis (H_a) is the claim that the researcher is trying to support. It is the conclusion that is accepted if the null hypothesis is rejected.
- The Claim: It generally states that a relationship exists or that a difference is present.
- The Outcome: Researchers seek to gather data that is so incompatible with the null hypothesis that they are forced to reject H_0 and accept H_a.

The Test and the Conclusion

- The hypothesis test uses sample data to calculate the p-value, which quantifies the probability of observing the data (or data more extreme) if the null hypothesis were true.
- Rejecting H_0: If the p-value is below a predetermined threshold (the significance level, usually 0.05), the data are deemed statistically incompatible with the null hypothesis, and the researcher rejects the null hypothesis.
- Failing to Reject H_0: If the p-value is high, the researcher simply fails to reject the null hypothesis, meaning the data do not provide sufficient evidence to support the alternative hypothesis. This does not mean the null hypothesis is proven true.

Significance Levels vs. P-Values

Significance levels and P-values are both central to hypothesis testing, but they serve different roles: the significance level is a fixed threshold set before the study, and the P-value is a calculated measure of evidence derived from the data.

Feature	**Significance Level (α)**	**P-Value**
When Set	Before the experiment	After the data analysis
Nature	A threshold for decision	A calculated probability of the data
Purpose	To define the acceptable risk of a Type I Error	To provide a measure of evidence against the Null Hypothesis

A P-value is often misunderstood; it does not measure the probability that the studied hypothesis is true or the probability that the data were produced by random chance alone. It also does not measure the size or importance of an effect.

Significance Level (α)

The significance level, denoted by α (alpha), is a pre-determined threshold that defines how much evidence is considered strong enough to reject the null hypothesis.

- Role: It is the standard for the test, representing the maximum risk of a Type I error (false positive) a researcher is willing to accept.
- Threshold: It is typically set at 0.05 (or 5%). This "bright-line" rule dictates the maximum probability of observing a result if the null hypothesis were true.
- Decision Rule: If the calculated P-value is less than or equal to α (P≤α), the result is deemed statistically significant, and the null hypothesis is rejected. A lower α (e.g., 0.01) demands stronger evidence.

P-Value

The P-value is a probability calculated from the observed data during the analysis.

- Role: Informally, it measures how incompatible the data are with the specified statistical model, often the null hypothesis.
- Definition: It is the probability of observing a result that is either equal to or more extreme than the actual observed result, assuming the null hypothesis is true.
- Interpretation: The smaller the P-value, the greater the statistical incompatibility of the data with the null hypothesis, suggesting stronger evidence against it.

Difference Between Type I and Type II Errors

The core difference between Type I and Type II errors lies in which hypothesis is incorrectly accepted or rejected during statistical hypothesis testing. In short, a Type I error is a false positive, and a Type II error is a false negative.

Feature	Type I Error	Type II Error
Error Type	False Positive	False Negative
Action	Reject H0 (Null Hypothesis)	Fail to Reject H0
Reality	H0 is True (No effect exists)	H0 is False (An effect exists)
Risk Controlled By	Significance Level (α)	Power (Related to β)

- H_0: Status Quo
- H_A: Statisitical Difference
- α: Significance Level (.05)
- p-value: Prob H_0 is true

Session 9: Regression

Nuts and Bolts of Regression

Core Purpose: Finding Relationships

Regression analysis is essentially the "detective work" that takes disorganized raw data and crafts it into meaningful conclusions.

- **The Goal:** To quantify the association between one or more independent variables (the explanatory factors) and a single dependent variable (the outcome you are trying to explain).
- **The Output:** A regression generates a formula that estimates the value of the dependent variable given the values of the independent variables.

Interpreting Regression

1. The Regression Coefficient

The most important part of the regression output is the regression coefficient.

- This is the estimated effect of a one-unit change in an independent variable on the dependent variable, while holding all other factors constant (the "ceteris paribus" condition).
- It allows a researcher to say, for example, "Every additional year of education is associated with an extra $1,000 in annual income, assuming everything else remains the same."

2. Statistical Significance

Regression results include a P-value for each coefficient, which indicates whether the observed relationship is likely due to chance. If the coefficient is statistically significant (typically $P < 0.05$), the researcher can be confident that the relationship is real and did not occur randomly.

3. $\mathbf{R^2}$ (R-squared)

The $\mathbf{R^2}$ statistic measures the overall explanatory power of the model. It represents the percentage of the variation in the dependent variable that is explained by all the independent variables included in the model. A higher R^2 suggests the model is a better fit for the data.

Key Pitfalls in Regression Analysis

- **Correlation vs. Causation:** The primary warning is that regression only finds correlation, not causation. While a model may show two variables move together, it cannot prove that one causes the other.
- **Controlling for Variables:** The selection of independent variables is crucial. Failing to include an important explanatory factor (omitted variable bias) can lead to a misleading conclusion about the variables that were included.
- **Multicollinearity:** This occurs when two independent variables are highly correlated with one another (e.g., IQ and education). This makes it difficult for the regression to isolate and determine the effect of a single variable, leading to unstable coefficients.
- **The Use of Proxies:** Sometimes, a researcher uses a substitute measure, or proxy variable, for the variable they are truly interested in, which can compromise the interpretation of the results.

Common Regression Mistakes

1. Correlation Does Not Equal Causation

This is the most critical mistake. Regression analysis is an excellent tool for identifying a correlation (an association) between variables, but it cannot prove causation. A statistically significant relationship only tells you that two things move together; it does not tell you why.

2. Reverse Causality

This error occurs when a researcher mistakenly assigns the role of cause to the effect and vice-versa.If A and B are correlated, it's easy to assume A causes B, when in fact, B may cause A.

3. Omitted Variable Bias (Confounding Variables)

This is a serious problem that occurs when a statistically important explanatory variable is left out of the model.

Leaving out a relevant variable not only makes the model less accurate overall but also leads to inaccurate and misleading estimates for the variables that were included, as they pick up the explanatory power of the missing factor.

4. Multicollinearity

Multicollinearity exists when two or more independent variables in the model are highly correlated with each other. This makes it mathematically difficult for the regression to isolate the independent effect of each of those variables. The result is that the estimated coefficients become unstable and their individual significance becomes questionable, even though the overall model's predictive power might still be high.

5. Using Regression to Analyze Non-Linear Relationships

Regression models taught at the basic level (Ordinary Least Squares, or OLS) are built on the assumption that the relationship between the variables is linear (a straight line). Applying a linear regression model to a relationship that is fundamentally non-linear (curved) will produce a formula that is an inaccurate fit and will not generalize well.

6. Extrapolating Beyond the Data

Regression models are only valid for a population that is similar to the sample on which the analysis was conducted. Using the regression formula to make predictions for groups that are outside the range of the original data (extrapolation) can lead to nonsensical results.

7. Data Mining (Using Too Many Variables)

Researchers can sometimes include a huge number of independent variables in a regression (or run a large number of separate regressions) until they find a few that are statistically significant. The risk is finding a spurious correlation that occurs purely by chance and has no theoretical basis or predictive power in the real world. This is often called p-hacking or data dredging.

Session 10: Investments in AI

Reflect on whether you consider the U.S. economy to be experiencing an "AI bubble" or if companies are on the verge of unlocking unlimited resources with AI (or, perhaps, somewhere in between).

- The U.S. economy is experiencing a dynamic that sits somewhere between a pure “AI bubble” and the verge of unlocking unlimited resources.
- The overall evidence strongly points toward a speculative boom that has outrun its fundamentals.
- The current situation displays several classic characteristics that suggest the market is overheated and experiencing a bubble, similar to the dot-com era.
- Despite lagging financial returns, the sheer scale of investment and the belief of major players suggest a genuine technological transformation is underway.
- The most accurate reflection is that the economy is currently stuck in the gap between massive investment and tangible realization.
- This gap aligns with Gartner’s prediction that AI is sliding toward the “trough of disillusionment”.

In what ways might an AI bubble today differ from the dot-com bubble of the 1990s?

- The potential AI bubble today differs from the dot-com bubble of the late 1990s in terms of systemic economic impact, the concentration of investment and revenue flow, and the nature of the financial dependencies.
- The AI investment frenzy presents a greater systemic risk to the overall economy than the dot-com bust.
- While both periods saw excessive hype, the funding structure and cash flow are distinct.
- Both bubbles feature a "productivity paradox," but the AI version is unique in its focus on generative technology.

What is the "productivity paradox?" Have you observed major efficiency gains from AI at your place of work?

The "productivity paradox" is a phenomenon where the introduction of a major new technology leads to huge investments by companies with scant evidence of a corresponding gain in workers' efficiency. It was first observed nearly four decades ago during the personal computer boom. Today, the same paradox is appearing with generative Artificial Intelligence (AI). The current situation is sometimes called the "gen. A.I. paradox".

Efficiency Gains: Processing data; Converting code between languages quickly; Write a starting point of code

The four red flags that investors are monitoring during this week's earnings releases. Make note of these.

- Disconnect Between Investment and Measurable Returns: There is a growing divergence between the massive, multibillion-dollar investments in AI infrastructure and the tangible, proven business outcomes.
- Circular Funding Deals: A growing web of interdependent AI deals is sparking memories of the 1990s dot-com bubble. These transactions often blur the line between buyer and seller.
- Surging AI Infrastructure Costs Dampening Profit Margins: While major tech companies (Microsoft, Alphabet, Amazon, and Meta) are expected to report solid revenue increases, surging capital expenditures related to AI infrastructure are dampening profit margins.
- Lagging Fundamentals vs. Investor Hype: The market rally is seen as running ahead of the actual business fundamentals. The high valuations are not yet supported by a solid business case for long-term growth.

Session 11: Perception and Design

How do we use principles of human perception to create high-quality visuals?

Reduce Cognitive Load (Eliminate Clutter)

Every element added to a visual takes up cognitive load (mental effort) on the part of the audience. Clutter—visual elements that take up space but don't increase understanding—causes excessive or extraneous cognitive load, which risks the audience ignoring the message.

Focus Attention (Preattentive Attributes)

Preattentive attributes are visual properties (like color, size, or shape) that our iconic memory picks up almost instantly, before conscious thought. Strategically using these elements is profoundly powerful because they direct the audience's attention exactly where you want it to focus. Preattentive attributes are used in two ways: (1) Direct Attention; (2) Create Visual Hierarchy.

Session 12: AI and Ethics -- AI and the Future Labor Force

AI and the future of work

IBM Article

The IBM article on AI and the future of work emphasizes that AI will fundamentally transform jobs by acting as a co-pilot rather than simply an automation replacement, necessitating a massive focus on reskilling the workforce.

The key insight is that the primary impact of AI will not be mass unemployment, but a shift toward "hybrid jobs" that require both technical aptitude and uniquely human skills like creativity, judgment, and emotional intelligence. IBM stresses that businesses must move beyond pilot projects to integrate AI across their entire enterprise, viewing it as an augmentation tool that frees workers from mundane tasks. Successfully navigating this transition requires organizations to foster a culture of continuous learning and invest heavily in training employees to collaborate effectively with AI, thus maximizing productivity and ensuring sustained competitiveness in the new era of work.

How Will AI Affect the Global Workforce?

Goldman Sachs Article

This Goldman Sachs research offers a surprisingly tempered view on the impact of AI on employment, suggesting the disruption will be modest and relatively temporary rather than leading to widespread job loss. While the firm estimates that 6-7% of the U.S. workforce, and up to 300 million jobs globally, could be displaced by generative AI, this is seen as a transitory impact. Economists predict that unemployment may only increase by about half a percentage point during the transition period as displaced workers seek new positions, with historical evidence suggesting such job displacement tends to disappear after approximately two years. The overall prediction is a net positive of new jobs emerging globally, especially for those who transition and upskill.

The displacement risk is highly concentrated in occupations performing routine cognitive tasks, such as computer programmers, accountants, administrative assistants, and customer service representatives. Conversely, AI is expected to boost labor productivity in developed markets by around 15% when fully adopted, which will increase corporate profits but may not benefit common employees directly, potentially increasing income inequality. To adapt, businesses must rethink their organizational structure, which is already showing signs of flattening as AI automates mid-level coordination tasks. The responsibility falls on both governments and employers to co-invest in upskilling programs and design new career pathways to ensure workers are prepared for this redefined future of work.

Session 14: How Will You Measure Your Life?

This epilogue, "How Will You Measure Your Life?", details the author's process for defining his life's purpose by focusing on three parts: sketching the person he wants to become, deeply committing to that likeness, and finding the right metric for measuring his life. The author initially drew his ideal "likeness" from his family values, faith, and profession, aspiring to be a man dedicated to improving the lives of others, characterized by kindness, honesty, and selflessness. The critical step was achieving a profound commitment to this likeness; he describes spending an hour every night reading, reflecting, and praying while at Oxford, knowing that this dedication to finding his purpose was more valuable than studying advanced academic subjects like econometrics.

The author realized, after a personal crisis regarding a church assignment, that while human organizations rely on aggregated metrics (like revenues, costs, or hierarchical titles) to understand the big picture, God's true measure of achievement is the individual. This led him to conclude that the only metrics that truly matter are the individuals he has helped to become better people. He reinforces this conclusion by recounting a period of despair after a stroke; he only recovered his sense of happiness and purpose when he stopped focusing on his own problems and resolved to focus his energy on helping resolve the challenges of others. He ultimately promises that figuring out one's purpose is the most important thing they will ever learn, trumping all business theories and academic knowledge.