AKSHAYA COLLEGE OF ENGINEERING AND TECHNOLOGY

PROJECT REPORT

"VARIOUS TECHNOLOGIES IN YEAR WISE"

SUBMITTED BY

SANNIBOYINA GUNA SRI TEJA

(720321104054)

UNDER THE GIUDANCE OF

1. ESTHER JEBA THANGAM.S

2. SARANYA NAIR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AKSHAYA COLLEGE OF ENGINEERING AND TECHNOLOGY

KINATHUKADAVU.

Report Submitted in Fulfilment of

Anna University in

R-PROGRAMMING (MINI PROJECT) LAB

(2023-2024)

“ANALYSIS OF VARIOUS TECHNOLOGIES”

ANALYZING THE EVOLUTION OF AI TRAINING AND COMPUTATION OVER THE PAST 3 DECADES

INTRODUCTION: Over the last thirty years, the field of Artificial Intelligence (AI) has undergone significant transformations, leading to groundbreaking advancements that have reshaped various industries and aspects of our daily lives. This project aims to provide a thorough analysis of the progression of AI training techniques and computational frameworks from the late 20th century to the present, emphasizing the crucial technological breakthroughs and trends that have propelled AI from theoretical concepts to practical applications.

By exploring the historical development of AI algorithms, hardware innovations, and key research contributions, this study seeks to unveil the narrative behind AI's journey from an emerging field to a pervasive and influential force in contemporary society. This analysis not only seeks to illuminate the past and present state of AI but also to offer insights into the future prospects and challenges that may arise in this dynamic and rapidly evolving domain. The outcomes of this research endeavor are anticipated to provide valuable insights for researchers, professionals, and enthusiasts aiming to comprehend the trajectory of AI evolution and the underlying forces driving its ongoing development.

AIM: This project aims to meticulously examine the evolution of AI training methodologies and computational frameworks over the past three decades, with the objective of comprehensively understanding the key technological advancements and trends that have shaped the contemporary landscape of AI research and applications. By analyzing historical milestones, significant algorithms, and hardware innovations, this study endeavors to unravel the intricate narrative of how AI has transitioned from theoretical concepts to practical implementations. Furthermore, the project seeks to offer insights into the potential future directions and challenges within the rapidly evolving AI domain. The ultimate goal is to provide a comprehensive overview that contributes to the collective understanding of AI's historical progression and its implications for future technological development and societal impact.

PLATFORM RStudio Version 4.3.1 is a powerful integrated development environment (IDE) designed specifically for R programming. Released as an upgrade to its predecessor, this version introduces several notable enhancements and features, aiming to streamline the R programming experience for both novice and experienced users. With an intuitive and user-friendly interface, RStudio 4.3.1 offers improved code navigation and debugging capabilities, empowering users to write, test, and debug R code seamlessly.

One of the key highlights of this version includes enhanced support for package development, facilitating efficient package creation and management within the R environment. Additionally, it provides improved integration with various data visualization libraries, allowing users to create compelling and insightful visual representations of their data. Furthermore, RStudio 4.3.1 incorporates advanced collaboration tools, enabling seamless collaboration and version control for team-based projects.

With its robust set of features and improved functionalities, RStudio Version 4.3.1 serves as a valuable tool for data scientists, researchers, and analysts, facilitating efficient and effective R programming for a wide range of data analysis tasks and research projects.

PROJECT SPECIFICATION: The project will involve a comprehensive analysis of the evolution of AI training methodologies and computational frameworks over the past three decades, focusing on key advancements and trends in the field. It will delve into historical milestones, prominent algorithms, and hardware innovations, aiming to uncover the transformative journey of AI from theoretical concepts to practical applications. The analysis will include a detailed exploration of the impact of AI on various industries and its potential future implications. The project will utilize RStudio Version 4.3.1 for data analysis, visualization, and code development, leveraging its advanced features for seamless collaboration and efficient package management. The final deliverable will consist of a detailed report, highlighting the significant findings and insights from the analysis, contributing to a comprehensive understanding of the historical progression and current state of AI technology.

HARDWARE SPECIFICATIONS:

Device specifications

Device name	DESKTOP-IME11DE
Processor	Intel(R) Core(TM) i3-4130 CPU @ 3.40GHz 3.40 GHz
Installed RAM	8.00 GB
Device ID	80E13BF3-461B-4ADB-9C5D-8960765CAF0A
Product ID	00331-90000-00001-AA762
System type	64-bit operating system, x64-based processor
Pen and touch	No pen or touch input is available for this display

DATASET:

Entity	Year	Day	Trainingcompute	Domain
6-layer MLP (MNIST)	2010	#	0.131	Vision
ADALINE	1960	#####	9.90E-12	Vision
ADAM (CIFAR-10)	2014	######	60.5	Vision
ALBERT-xxlarge	2020	#####	2390000	Language
ALVINN	1989	#####	0.0000812	Driving
AlexNet	2012	######	470	Vision
AlexaTM 20B	2022	####	204000000	Language
AlphaCode	2022	####	405000000	Language
AlphaFold	2020	#####	100000	Other
AlphaGo Fan	2015	####	380000.03	Games
AlphaGo Lee	2016	###	1900000	Games
AlphaGo Master	2017	####	150000000	Games
AlphaGo Zero	2017	####	341000000	Games
AlphaStar	2019	#####	202000020	Games
AlphaZero	2017	#####	36700000	Games
AmoebaNet-A (F=448)	2018	####	385000	Vision
BERT-Large	2018	#####	285000	Language
BLOOM	2022	#####	180000000	Language
Back-propagation	1986	######	1.24E-07	Other
BiLSTM for Speech	2005	#####	0.024100002	Speech
BigGAN-deep 512x512	2018	#####	3000000	Drawing
Chinchilla	2022	######	576000000	Language
ConSERT	2021	####	280000	Language
DALL-E	2021	#####	47000000	Drawing
DQN	2013	######	2.3000002	Games
Decision tree (classification)	2001	#####	0.063	Vision
Decoupled weight decay regularization	2019	#####	2470.0002	Vision
DeepSpeech2 (English)	2015	###	26000	Speech
Denoising Diffusion	2021	####	95000.01	Drawing
Dropout (CIFAR)	2012	####	4.27	Vision
Dropout (ImageNet)	2012	###	273	Vision
Dropout (MNIST)	2012	####	6.0400004	Vision

EfficientNetV2	2021	#####	6220	Vision	
FLAN	2021	######	49000004	Language	
Falcon 180B	2023	######	3780000000	Language	
Feedforward NN	2010	######	0.35000002	Vision	
Flamingo	2022	######	270000000	Multimodal	
Fuzzy NN	1992	######	0.0000014	Speech	
GANs	2014	######	518	Drawing	
GNMT	2016	######	6900000	Language	
GPT	2018	######	17600	Language	
GPT-2 (1542M)	2019	######	1490000	Language	
GPT-3 175B (davinci)	2020	######	314000000	Language	
GPT-4	2023	######	21000000000	Multimodal	
GPT-NeoX-20B	2022	######	93200010	Language	
Gato	2022	######	5440000	Multimodal	
GoogLeNet / InceptionV1	2015	#######	1560	Vision	
Hanabi 4 player	2019	#####	4300	Games	
HuBERT	2021	######	5540000	Language	
Image Classification	2013	######	0.090799995	Vision	
Image generation	2013	######	0.475	Vision	
Innervator	1989	######	1.20E-07	Other	
JFT	2017	######	479000	Vision	
Jais	2023	######	30800002	Language	
KN5 LM + RNN 400/10	2010	######	61.4	Speech	
KataGo	2019	######	23200.002	Games	
LLaMA-65B	2023	######	550000000	Language	
LSTM	1997	######	0.021	Language	
LaMDA	2022	######	355000000	Language	
LeNet-5	1998	######	0.00281	Vision	
Libratus	2017	######	551000	Games	
Llama 2	2023	######	900000000	Language	
M6-T	2021	######	5500000.5	Multimodal	
MCDNN (MNIST)	2012	#######	3.73	Vision	
MSRA (C, PReLU)	2015	######	24000	Vision	
Megatron-Turing	2021	#######	1170000000	Language	
Meta Pseudo Labels	2021	######	47900004	Vision	
Minerva (540B)	2022	######	2740000300	Language	
Mitosis	2013	######	137	Vision	
MnasNet-A1 + SSDLite	2019	######	1500000	Vision	
MnasNet-A3	2019	######	1500000	Vision	
MuZero	2019	######	48000	Games	
NASv3 (CIFAR-10)	2016	######	2200000	Vision	
NLLB	2022	#####	17500000	Language	
NPLM	2003	######	1.3000001	Language	
Named Entity	2016	######	96.9	Language	

Neocognitron	1980	#######	2.28E-07	Vision	
NetTalk	1987	#######	2.77E-05	Speech	
NÃœWA	2021	#######	4840000	Multimodal	
ObjectNet	2019	#######	19400.002	Vision	
OpenAI Five	2019	#######	67000004	Games	
OpenAI Five Rerun	2019	#######	13000000	Games	
OpenAI TI7 DOTA 1v1	2017	#######	604609.5	Games	
PLUG	2021	######	36000000	Language	
PNASNet-5	2017	######	66300		
PaLM (540B)	2022	#####	2530000100	Language	
PaLM 2	2023	######	7339999700	Language	
Pandemonium (morse)	1959	#######	6.00E-07	Other	
Part-of-sentence tagging model	2016	#######	145	Language	
Parti	2022	#######	396000000	Drawing	
Perceptron Mark I	1957	#######	6.95E-10	Vision	
R-FCN	2016	#######	61.500004	Vision	
RNN 500/10 + RT09 LM	2010	######	3.41	Speech	
RNNsearch-50*	2014	######	1560	Language	
ResNet-152 (ImageNet)	2015	######	12100	Vision	
RetinaNet-R101	2017	######	2070	Vision	
RoBERTa Large	2019	######	4150000	Language	
SPPNet	2014	######	3410	Vision	
Samuel Neural Checkers	1959	######	4.28E-07	Games	
Seq2Seq LSTM	2014	######	56000	Language	
SmooCT	2014	######	69	Games	
Sparse all-MLP	2022	######	60800	Language	
Stable Diffusion	2022	######	50000000	Drawing	
Swift	2023	######	53.37	Robotics	
System 11	1996	######	0.0000129	Vision	
T5-11B	2019	######	40500000	Language	
T5-3B	2019	######	10400001	Language	
TD-Gammon	1992	######	0.018199999	Games	
Taiyi-Stable Diffusion	2022	######	51000000	Drawing	
Theseus	1950	######	4.00E-14	Other	
TransE	2013	######	1340	Other	
Transformer	2017	######	7420	Language	
Transformer local	2021	######	24100	Vision	
Unsupervised High-level	2012	######	600	Vision	
VGG16	2014	######	9250	Vision	
ViT-H/14	2020	#######	12800000	Vision	
Visualizing CNNs	2013	#######	532	Vision	
Whisper	2022	#######	46500000	Speech	
Word2Vec (large)	2013	#######	38.9	Language	
YOLOv3	2018	######	50900	Vision	

Zip CNN		1989	######	0.0000434	Vision	
wave2vec 2.0 LARGE		2020	######	1900000	Speech	

PACKAGES REQUIRED:

1) FORCASTE:

The "forecast" package in R language serves as a comprehensive and powerful tool for time series forecasting and prediction analysis. Developed by Rob J Hyndman and his team, this package provides a wide array of sophisticated algorithms and functions for effectively analyzing and forecasting time series data. With its intuitive interface and extensive documentation, the "forecast" package offers users the flexibility to apply various forecasting methods, including exponential smoothing, ARIMA, and state space models, among others.

The package facilitates the seamless integration of advanced time series analysis techniques, enabling users to handle complex data patterns and fluctuations with ease. Moreover, it offers robust functionalities for visualizing and interpreting forecast results, empowering users to make informed decisions based on accurate and reliable predictions. The "forecast" package has become an essential resource for researchers, analysts, and data scientists seeking to gain deeper insights into time series data and make informed forecasts for a wide range of applications across diverse industries.

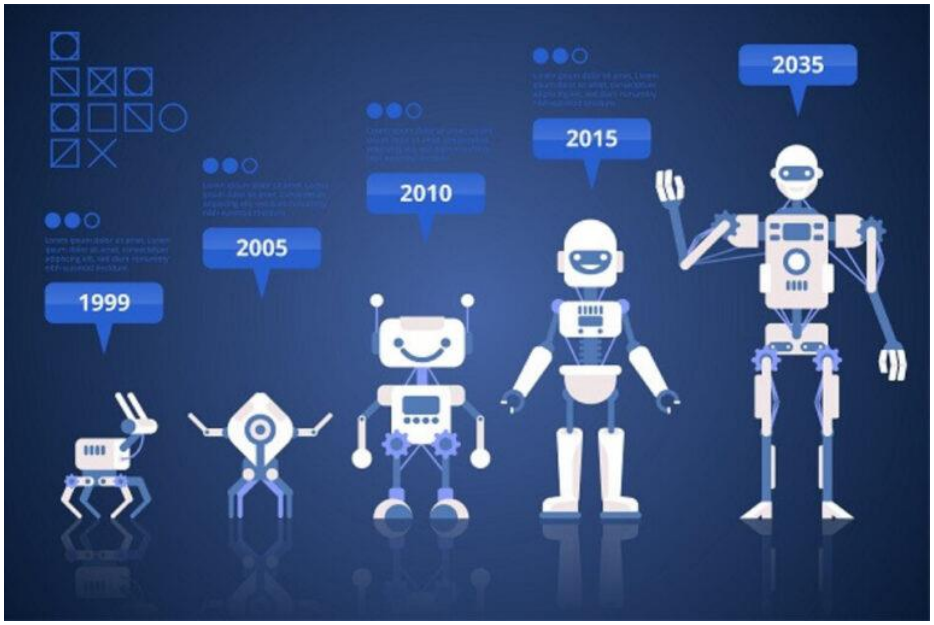
2) GGPLOT2:

GGPLOT2 is a widely used data visualization package in R, known for its powerful capabilities in creating visually appealing and customizable plots. Built on the grammar of graphics concept, GGPLOT2 allows users to construct complex and informative data visualizations with relative ease. Its intuitive syntax and extensive options for customization enable users to generate a wide range of plots, including scatter plots, bar plots, histograms, box plots, and more. GGPLOT2's versatility and flexibility make it a popular choice for data analysts and researchers seeking to communicate insights effectively through compelling and publication-quality visualizations.

CHAPTER 1

- **WHAT IS ANALYZING THE EVOLUTION OF AI TRAINING AND COMPUTATION?**

Analyzing the evolution of AI training and computation involves a comprehensive examination of the historical progression, advancements, and trends in the methodologies and computational frameworks used in the development of Artificial Intelligence (AI) systems. This process entails studying the transformative journey of AI from its early theoretical foundations to its current practical applications across various domains. It includes a meticulous investigation of the development of AI algorithms, hardware infrastructure, and computational models, tracing the key milestones and breakthroughs that have shaped the field's evolution. Moreover, this analysis entails a deep understanding of the changing paradigms in AI training, encompassing the shifts in data processing techniques, machine learning algorithms, and neural network architectures over the years. By scrutinizing the evolution of AI training and computation, researchers and practitioners gain valuable insights into the underlying factors driving the progress of AI technology, enabling them to anticipate future trends, challenges, and opportunities in this dynamic and rapidly evolving field.



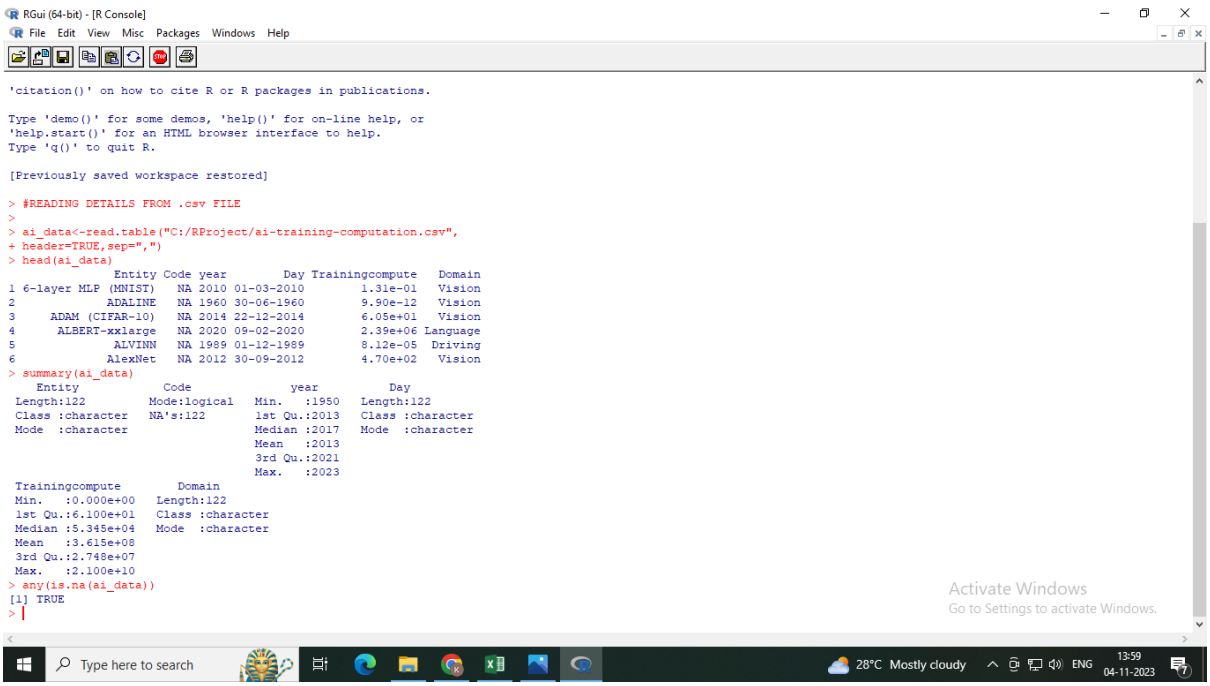
IMPLEMENTATION:

1) READING DETAILS FROM .CSV FILE:

CODE:

```
#READING DETAILS FROM .csv FILE
ai_data<-read.table("C:/RProject/ai-training-computation.csv",
header=TRUE,sep=",")
head(ai_data)
summary(ai_data)
```

OUTPUT:



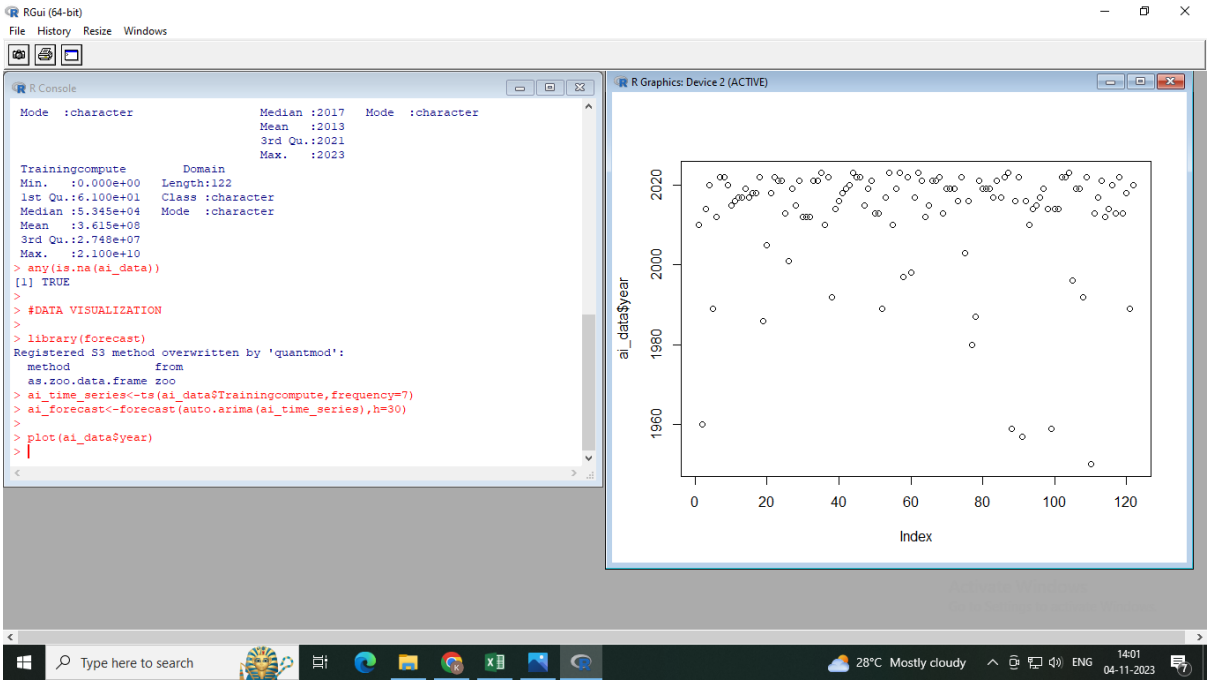
CHAPTER 2

1) VISUALIZATION PLOT

CODE:

```
#DATA VISUALIZATION
library(forecast)
ai_time_series<-ts(ai_data$Trainingcompute,frequency=7)
ai_forecast<-forecast(auto.arima(ai_time_series),h=30)
plot(ai_data$year)
```

OUTPUT:

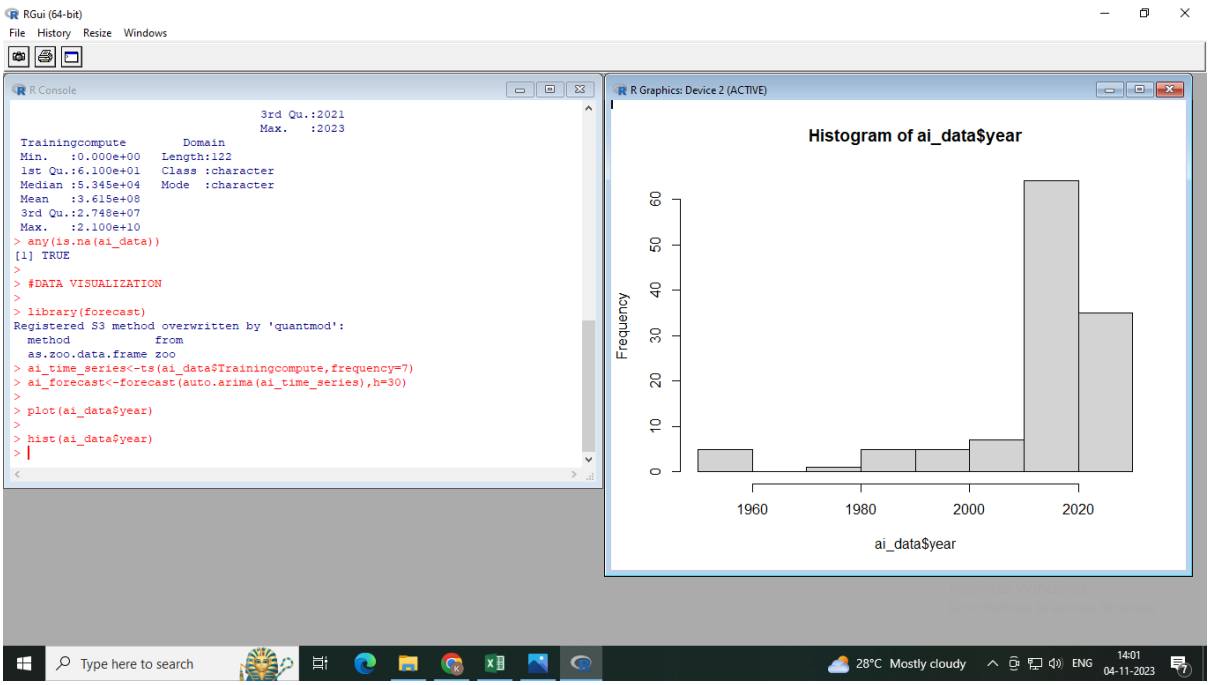


2) **HISTOGRAM**

CODE:

```
hist(ai_data$year)
```

OUTPUT:

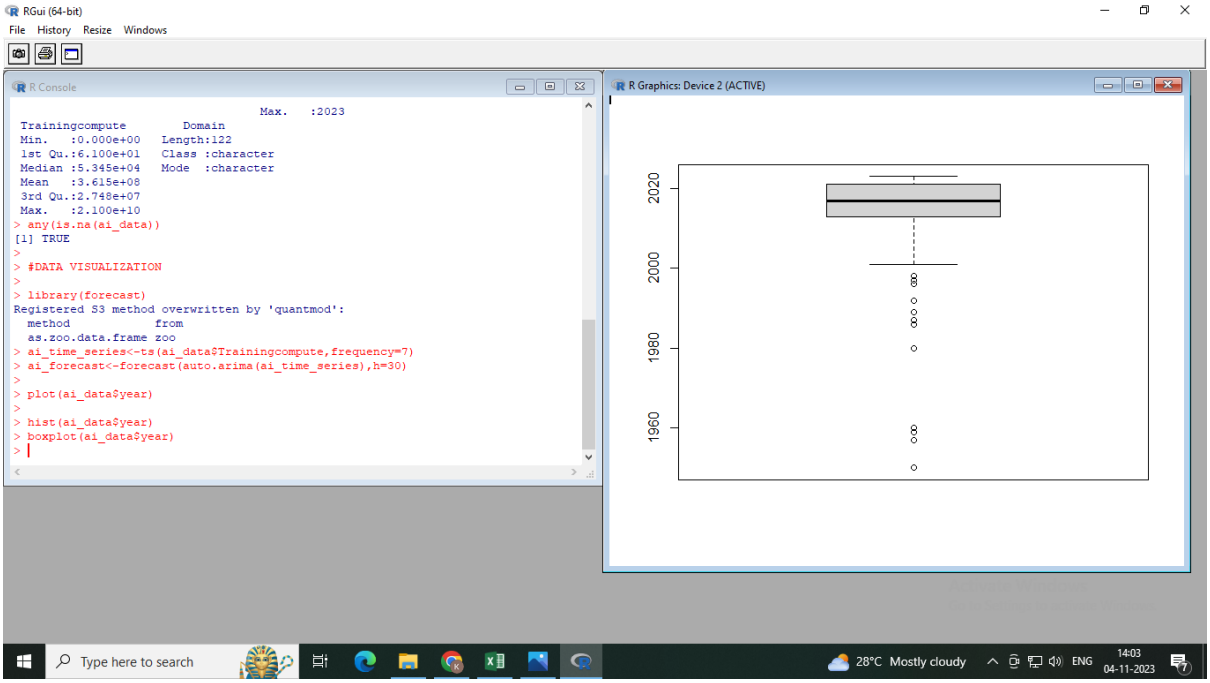


3) BOXPLOT

CODE:

```
boxplot(ai_data$year)
```

OUTPUT:



CHAPTER 3

CASESTUDY:

Analyzing the evolution of AI training and computation involves a comprehensive examination of the historical, technological, and methodological advancements in the field of Artificial Intelligence (AI) over a significant period. It encompasses a detailed investigation of the various approaches, algorithms, and computational frameworks that have shaped the development and application of AI across different domains. The analysis aims to elucidate the progression of AI from its early conceptualization to its current sophisticated implementations, highlighting the transformative impact of technological breakthroughs on the capabilities and efficiency of AI systems. This examination involves the study of foundational concepts such as machine learning, deep learning, and neural networks, tracing

their evolution and refinement over time. Additionally, it includes an exploration of the hardware and software innovations that have facilitated the growth and deployment of AI technologies, leading to increased computational power, efficiency, and scalability. By scrutinizing the historical context and contemporary advancements, this analysis aims to provide insights into the key drivers, challenges, and trends that have shaped the landscape of AI training and computation, ultimately contributing to a comprehensive understanding of the past, present, and future trajectory of AI development and its implications for various industries and societal domains.

KMEANS ALGORITHEM:

K-means Algorithm: Unveiling Patterns and Clusters in Data

The K-means algorithm serves as a fundamental tool in the realm of unsupervised machine learning, designed to uncover underlying patterns and structures within a dataset. This iterative clustering technique aims to partition a given dataset into K clusters, where each data point is assigned to the cluster with the nearest mean. The algorithm's objective is to minimize the within-cluster variance, ensuring that the data points within each cluster are as similar to each other as possible.

The fundamental steps of the K-means algorithm are as follows:

1) Initialization: Begin by selecting K initial centroids, either randomly or through specific initialization techniques. These centroids serve as the initial cluster centers.

2) Assignment: Assign each data point to the nearest centroid, forming K clusters based on the minimum Euclidean distance from each centroid.

3)Update: Recalculate the centroids of the clusters based on the mean of the data points within each cluster.

4)Iterate: Repeat the assignment and update steps until the centroids no longer change significantly or until a predetermined number of iterations is reached.

The algorithm's performance is heavily dependent on the initial choice of centroids, as it may converge to local optima rather than the global optimum. Various techniques, such as the K-means++ algorithm, have been proposed to improve the selection of initial centroids and enhance the convergence speed.

While K-means is a widely used clustering algorithm, it is important to note its limitations, including its sensitivity to outliers, dependence on the initial number of clusters (K), and its assumption that the clusters are spherical and of equal size. Additionally, the algorithm may struggle with non-linear data and high-dimensional datasets, often requiring preprocessing techniques or the use of dimensionality reduction methods.

CODE:

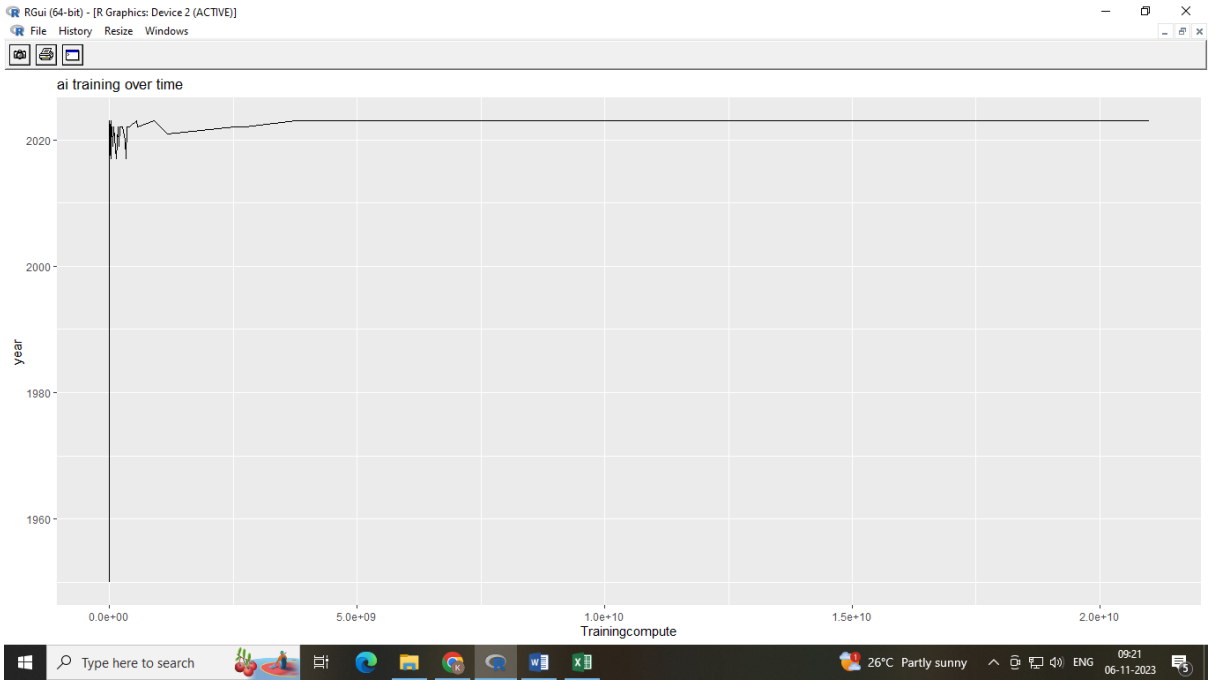
```
# Sample data for demonstration purposes

set.seed(123)

x <- c(rnorm(50, 0, 3), rnorm(50, 15, 3))

y <- c(rnorm(50, 0, 3), rnorm(50, 15, 3))

data <- data.frame(x, y)

# Implementing K-means algorithm

k <- 2  # Number of clusters
```

```
kmeans_result <- kmeans(data, centers = k)

# Displaying the cluster centers

cat("Cluster Centers:\n")

print(kmeans_result$centers)

# Displaying the cluster assignments

cat("\nCluster Assignments:\n")

print(kmeans_result$cluster)
```

CHAPTER 4

VISUALIZATION:

```
> library(ggplot2)
> ggplot(ai_data,aes(Trainingcompute,year))+geom_line()+
+ labs(title="ai training over time",x="Trainingcompute",y="year")
>
```



FULL CODING

#READING DETAILS FROM .csv FILE

ai_data<-read.table("C:/RProject/ai-training-computation.csv",

header=TRUE,sep=",")

head(ai_data)

summary(ai_data)

any(is.na(ai_data))

#DATA VISUALIZATION

library(forecast)

ai_time_series<-ts(ai_data$Trainingcompute,frequency=7)

ai_forecast<-forecast(auto.arima(ai_time_series),h=30)

plot(ai_data$year)

hist(ai_data$year)

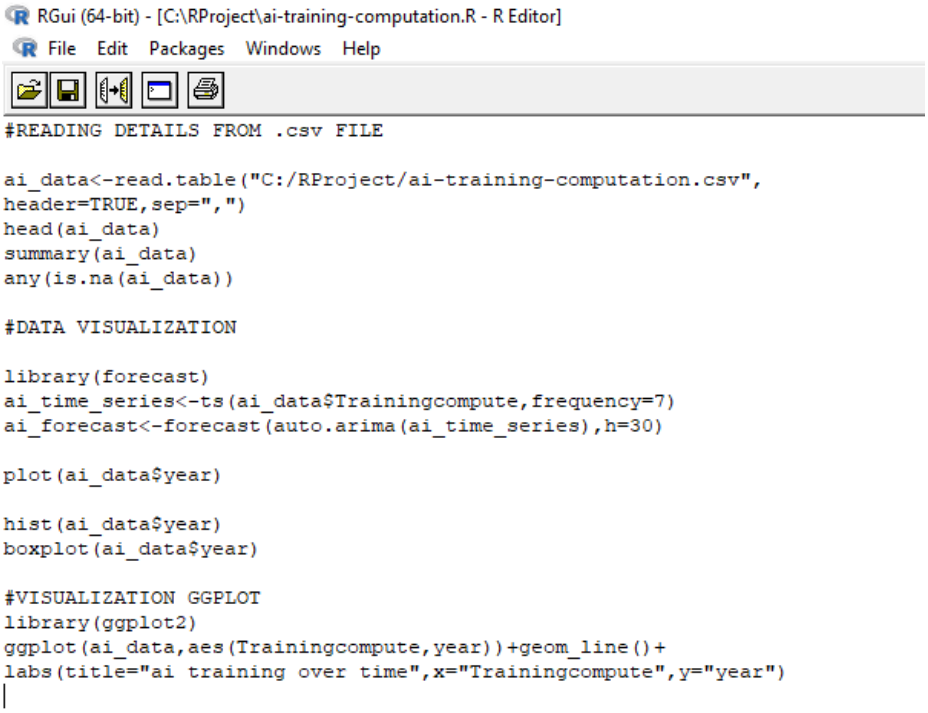
boxplot(ai_data$year)

#VISUALIZATION GGPLOT

library(ggplot2)

ggplot(ai_data,aes(Trainingcompute,year))+geom_line()+

labs(title="ai training over time",x="Trainingcompute",y="year")



CONCLUSION:

Unveiling the Trajectory of AI Evolution

Through a meticulous analysis of the evolution of AI training and computation over the past three decades, this project has shed light on the remarkable advancements and transformative developments that have shaped the landscape of Artificial Intelligence. By delving into the historical progression of AI training methodologies, computational frameworks, and technological innovations, we have gained a comprehensive understanding of the pivotal milestones that have propelled AI from its conceptual origins to its current pervasive presence in various industries and societal domains.

The examination of foundational concepts such as machine learning, deep learning, and neural networks has highlighted the progressive refinement and diversification of AI applications, underscoring its potential to drive innovation and address complex real-world challenges. Moreover, the exploration of hardware advancements and software frameworks has emphasized the crucial role of computational power and infrastructure in facilitating the accelerated growth and deployment of AI technologies.

As AI continues to revolutionize numerous sectors, including healthcare, finance, manufacturing, and beyond, it is imperative to recognize the ethical implications and societal impact associated with its widespread adoption. Understanding the historical context and current state of AI technology is pivotal in fostering responsible and ethical AI development, ensuring that its benefits are harnessed while mitigating potential risks and challenges.

This project not only contributes to the collective understanding of AI's evolution but also serves as a foundation for future research endeavors and innovations in the dynamic and rapidly evolving field of Artificial Intelligence. It is our hope that the insights gleaned from this analysis will foster continued progress and responsible utilization of AI for the betterment of society.

FUTURE ENHANCEMENT:

Moving forward, this project can be further augmented through the integration of advanced machine learning models and deep learning architectures, enabling a more nuanced analysis of complex data patterns and facilitating predictive modeling in diverse domains. Embracing cutting-edge AI technologies, such as reinforcement learning and generative adversarial networks (GANs), could offer valuable insights into dynamic decision-making processes and the generation of synthetic data for enhanced training and simulation purposes. Additionally, exploring the integration of AI-driven natural language processing (NLP) and computer vision algorithms could open new avenues for comprehensive data understanding and interpretation, enabling the extraction of meaningful insights from unstructured data sources and multimedia content. Leveraging distributed computing frameworks and cloud-based infrastructure can further amplify the project's scalability and performance, enabling the analysis of massive datasets and real-time processing for time-sensitive applications. Moreover, considering the ethical implications and societal impact of AI technologies in

future research endeavors remains crucial, emphasizing the need for responsible AI development and the implementation of robust governance frameworks to ensure the ethical and equitable deployment of AI solutions across diverse socio-economic contexts.