**CS455 Assignment 1**

**Matt Reeder, Tara Williams, and Lucas Smith**

**Data Description**

For this assignment, we pulled airline data for the year 1987 from the website <http://stat-computing.org/dataexpo/2009/the-data.html> . This data contains 1,311,827 entries of flight data. We downloaded the CSV file directly, then used a Python script to pull the data from the CSV, randomly select 10,000 entries, and input them into a new CSV. Since the data had so many attributes, we stripped out all attributes but the five we needed when we put the data in the new CSV file. From the original CSV file, we took two numerical attributes, one binary attribute, and two categorical attributes. Our program is run by navigating in the terminal to the folder where **main.py** is stored and typing **python main.py**. This will run the program, create a new CSV file, and output the computed statistics for that CSV file.

**Actual Elapsed Time:** This is the time it takes for the airplane to reach its destination. This is a numerical attribute, represented as a floating point number.

**Distance:** This is the distance the plane travels on its journey from one airport to another. This is a numerical attribute, also represented as a floating point number.

**Cancelled:** This is a binary attribute that determines whether or not the plane’s flight was cancelled. A cancelled flight is represented by a 1, and a non-cancelled flight is represented by a 0.

**Origin:** This is the airport where the plane takes off on its journey. This is a categorical attribute represented by a string.

**Destination:** This is the airport where the plane lands at the end of its journey. This is a categorical attribute represented by a string.

**Data Statistics**

To compute the statistics for this data, we wrote a Python script that reads the columns from our downloaded file, which is named **1987.csv**. Then, we wrote methods to compute the minimum, maximum, mean, median, and standard deviation for numerical data and methods to compute the frequencies for categorical/nominal and binary data. We used the file **example\_stats.csv** to get the data below. This is included in our project folder and will remain unchanged; it is a static example of one run of our program. Our program downloads different rows from the original CSV file every time we run it, and this is stored in our new file called **downloaded\_stats.csv**.

**Actual Elapsed Time:**

Min: 13.0

Max: 540.0

Mean: 101.204

Median: 94.0

Standard Deviation: 60.454

**Distance:**

Min: 11.0

Max: 4243.0

Mean: 583.582

Median: 1452.0

Standard Deviation: 485.018

**Cancelled:**

Cancelled: True Frequency: 131 Percentage: 1.31%

Cancelled: False Frequency: 9869 Percentage: 98.69%

**Origin:**

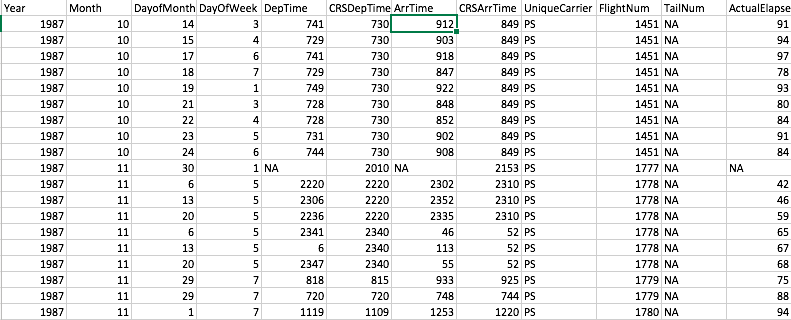
|  |  |
| --- | --- |
| Origin: JFK Frequency: 98 Percentage: 0.98 %  Origin: GSP Frequency: 16 Percentage: 0.16 %  Origin: FNT Frequency: 1 Percentage: 0.01 %  Origin: SIT Frequency: 3 Percentage: 0.03 %  Origin: MIA Frequency: 140 Percentage: 1.4 %  Origin: BOS Frequency: 211 Percentage: 2.11 %  Origin: OAK Frequency: 68 Percentage: 0.68 %  Origin: FWA Frequency: 13 Percentage: 0.13 %  Origin: CCR Frequency: 4 Percentage: 0.04 %  Origin: LIT Frequency: 32 Percentage: 0.32 %  Origin: BOI Frequency: 15 Percentage: 0.15 %  Origin: BGR Frequency: 2 Percentage: 0.02 %  Origin: YUM Frequency: 3 Percentage: 0.03 %  Origin: DRO Frequency: 3 Percentage: 0.03 %  Origin: CPR Frequency: 8 Percentage: 0.08 %  Origin: AGS Frequency: 9 Percentage: 0.09 %  Origin: GSO Frequency: 46 Percentage: 0.46 %  Origin: SAN Frequency: 112 Percentage: 1.12 %  Origin: PIA Frequency: 8 Percentage: 0.08 %  Origin: MYR Frequency: 10 Percentage: 0.1 %  Origin: PIE Frequency: 2 Percentage: 0.02 %  Origin: DCA Frequency: 170 Percentage: 1.7 %  Origin: LBB Frequency: 18 Percentage: 0.18 %  Origin: HSV Frequency: 9 Percentage: 0.09 %  Origin: RAP Frequency: 10 Percentage: 0.1 %  Origin: BWI Frequency: 116 Percentage: 1.16 %  Origin: PIT Frequency: 208 Percentage: 2.08 %  Origin: RIC Frequency: 34 Percentage: 0.34 %  Origin: AMA Frequency: 19 Percentage: 0.19 %  Origin: SAT Frequency: 57 Percentage: 0.57 %  Origin: CHA Frequency: 12 Percentage: 0.12 %  Origin: ILM Frequency: 12 Percentage: 0.12 %  Origin: IAH Frequency: 153 Percentage: 1.53 %  Origin: TPA Frequency: 122 Percentage: 1.22 %  Origin: ALO Frequency: 4 Percentage: 0.04 %  Origin: IAD Frequency: 103 Percentage: 1.03 %  Origin: BGM Frequency: 4 Percentage: 0.04 %  Origin: JAN Frequency: 19 Percentage: 0.19 %  Origin: HRL Frequency: 11 Percentage: 0.11 %  Origin: CHS Frequency: 32 Percentage: 0.32 %  Origin: EYW Frequency: 5 Percentage: 0.05 %  Origin: GPT Frequency: 1 Percentage: 0.01 %  Origin: BNA Frequency: 97 Percentage: 0.97 %  Origin: SPN Frequency: 6 Percentage: 0.06 %  Origin: JAC Frequency: 1 Percentage: 0.01 %  Origin: LAN Frequency: 7 Percentage: 0.07 %  Origin: OME Frequency: 1 Percentage: 0.01 %  Origin: JNU Frequency: 8 Percentage: 0.08 %  Origin: SNA Frequency: 56 Percentage: 0.56 %  Origin: PHL Frequency: 132 Percentage: 1.32 %  Origin: PHF Frequency: 2 Percentage: 0.02 %  Origin: BIS Frequency: 9 Percentage: 0.09 %  Origin: SYR Frequency: 71 Percentage: 0.71 %  Origin: PHX Frequency: 201 Percentage: 2.01 %  Origin: LAX Frequency: 331 Percentage: 3.31 %  Origin: MAF Frequency: 14 Percentage: 0.14 %  Origin: MBS Frequency: 7 Percentage: 0.07 %  Origin: APF Frequency: 2 Percentage: 0.02 %  Origin: LAS Frequency: 141 Percentage: 1.41 %  Origin: GUC Frequency: 1 Percentage: 0.01 %  Origin: CRP Frequency: 8 Percentage: 0.08 %  Origin: CRW Frequency: 12 Percentage: 0.12 %  Origin: CMI Frequency: 6 Percentage: 0.06 %  Origin: CMH Frequency: 47 Percentage: 0.47 %  Origin: LMT Frequency: 2 Percentage: 0.02 %  Origin: GJT Frequency: 10 Percentage: 0.1 %  Origin: GUM Frequency: 3 Percentage: 0.03 %  Origin: FLL Frequency: 65 Percentage: 0.65 %  Origin: DEN Frequency: 318 Percentage: 3.18 %  Origin: BZN Frequency: 8 Percentage: 0.08 %  Origin: DTW Frequency: 199 Percentage: 1.99 %  Origin: EVV Frequency: 4 Percentage: 0.04 %  Origin: LSE Frequency: 3 Percentage: 0.03 %  Origin: MFE Frequency: 1 Percentage: 0.01 %  Origin: SCK Frequency: 5 Percentage: 0.05 %  Origin: SFO Frequency: 262 Percentage: 2.62 %  Origin: BUR Frequency: 58 Percentage: 0.58 %  Origin: ROA Frequency: 7 Percentage: 0.07 %  Origin: ROC Frequency: 45 Percentage: 0.45 %  Origin: MFR Frequency: 6 Percentage: 0.06 %  Origin: GEG Frequency: 26 Percentage: 0.26 %  Origin: BTM Frequency: 2 Percentage: 0.02 %  Origin: IDA Frequency: 1 Percentage: 0.01 %  Origin: BUF Frequency: 55 Percentage: 0.55 %  Origin: ANC Frequency: 26 Percentage: 0.26 %  Origin: TRI Frequency: 2 Percentage: 0.02 %  Origin: GRR Frequency: 22 Percentage: 0.22 %  Origin: SHV Frequency: 12 Percentage: 0.12 %  Origin: BDL Frequency: 62 Percentage: 0.62 %  Origin: CSG Frequency: 4 Percentage: 0.04 %  Origin: DLH Frequency: 2 Percentage: 0.02 %  Origin: YKM Frequency: 1 Percentage: 0.01 %  Origin: ABE Frequency: 9 Percentage: 0.09 %  Origin: ROR Frequency: 2 Percentage: 0.02 %  Origin: STL Frequency: 265 Percentage: 2.65 %  Origin: MLI Frequency: 17 Percentage: 0.17 %  Origin: EUG Frequency: 10 Percentage: 0.1 %  Origin: STT Frequency: 7 Percentage: 0.07 %  Origin: ABQ Frequency: 61 Percentage: 0.61 %  Origin: KTN Frequency: 5 Percentage: 0.05 %  Origin: HOU Frequency: 98 Percentage: 0.98 %  Origin: HPN Frequency: 4 Percentage: 0.04 %  Origin: MLU Frequency: 8 Percentage: 0.08 %  Origin: CWA Frequency: 4 Percentage: 0.04 %  Origin: CHO Frequency: 2 Percentage: 0.02 %  Origin: STX Frequency: 6 Percentage: 0.06 %  Origin: ICT Frequency: 26 Percentage: 0.26 %  Origin: MCO Frequency: 137 Percentage: 1.37 %  Origin: PWM Frequency: 7 Percentage: 0.07 %  Origin: BHM Frequency: 43 Percentage: 0.43 %  Origin: VPS Frequency: 1 Percentage: 0.01 %  Origin: RSW Frequency: 24 Percentage: 0.24 % | Origin: GCN Frequency: 4 Percentage: 0.04 %  Origin: ORF Frequency: 53 Percentage: 0.53 %  Origin: DSM Frequency: 28 Percentage: 0.28 %  Origin: GRB Frequency: 8 Percentage: 0.08 %  Origin: EWR Frequency: 222 Percentage: 2.22 %  Origin: MHT Frequency: 5 Percentage: 0.05 %  Origin: PBI Frequency: 44 Percentage: 0.44 %  Origin: BTV Frequency: 10 Percentage: 0.1 %  Origin: RNO Frequency: 45 Percentage: 0.45 %  Origin: BTR Frequency: 26 Percentage: 0.26 %  Origin: PSC Frequency: 7 Percentage: 0.07 %  Origin: AVP Frequency: 7 Percentage: 0.07 %  Origin: PSG Frequency: 2 Percentage: 0.02 %  Origin: MGM Frequency: 6 Percentage: 0.06 %  Origin: MOT Frequency: 10 Percentage: 0.1 %  Origin: AVL Frequency: 1 Percentage: 0.01 %  Origin: HLN Frequency: 1 Percentage: 0.01 %  Origin: OKC Frequency: 63 Percentage: 0.63 %  Origin: HTS Frequency: 5 Percentage: 0.05 %  Origin: IND Frequency: 64 Percentage: 0.64 %  Origin: MOB Frequency: 12 Percentage: 0.12 %  Origin: PSP Frequency: 6 Percentage: 0.06 %  Origin: TYS Frequency: 23 Percentage: 0.23 %  Origin: ILG Frequency: 1 Percentage: 0.01 %  Origin: COS Frequency: 18 Percentage: 0.18 %  Origin: FSD Frequency: 10 Percentage: 0.1 %  Origin: ATL Frequency: 519 Percentage: 5.19 %  Origin: ISP Frequency: 17 Percentage: 0.17 %  Origin: HNL Frequency: 25 Percentage: 0.25 %  Origin: ATW Frequency: 2 Percentage: 0.02 %  Origin: ISO Frequency: 1 Percentage: 0.01 %  Origin: ACV Frequency: 5 Percentage: 0.05 %  Origin: SMF Frequency: 56 Percentage: 0.56 %  Origin: SUX Frequency: 6 Percentage: 0.06 %  Origin: FCA Frequency: 2 Percentage: 0.02 %  Origin: CID Frequency: 8 Percentage: 0.08 %  Origin: TVL Frequency: 3 Percentage: 0.03 %  Origin: PVD Frequency: 25 Percentage: 0.25 %  Origin: SEA Frequency: 131 Percentage: 1.31 %  Origin: SLC Frequency: 112 Percentage: 1.12 %  Origin: MDW Frequency: 33 Percentage: 0.33 %  Origin: MDT Frequency: 13 Percentage: 0.13 %  Origin: RDU Frequency: 100 Percentage: 1.0 %  Origin: LNK Frequency: 14 Percentage: 0.14 %  Origin: GTF Frequency: 4 Percentage: 0.04 %  Origin: PDX Frequency: 79 Percentage: 0.79 %  Origin: CLE Frequency: 123 Percentage: 1.23 %  Origin: DFW Frequency: 429 Percentage: 4.29 %  Origin: LFT Frequency: 3 Percentage: 0.03 %  Origin: SJU Frequency: 23 Percentage: 0.23 %  Origin: RDD Frequency: 1 Percentage: 0.01 %  Origin: AUS Frequency: 66 Percentage: 0.66 %  Origin: SRQ Frequency: 12 Percentage: 0.12 %  Origin: YAK Frequency: 1 Percentage: 0.01 %  Origin: CLT Frequency: 188 Percentage: 1.88 %  Origin: SJC Frequency: 85 Percentage: 0.85 %  Origin: ELP Frequency: 42 Percentage: 0.42 %  Origin: TLH Frequency: 13 Percentage: 0.13 %  Origin: MSO Frequency: 9 Percentage: 0.09 %  Origin: OMA Frequency: 36 Percentage: 0.36 %  Origin: BLI Frequency: 2 Percentage: 0.02 %  Origin: ITH Frequency: 2 Percentage: 0.02 %  Origin: BET Frequency: 1 Percentage: 0.01 %  Origin: MEM Frequency: 138 Percentage: 1.38 %  Origin: LYH Frequency: 3 Percentage: 0.03 %  Origin: TUS Frequency: 42 Percentage: 0.42 %  Origin: ALB Frequency: 27 Percentage: 0.27 %  Origin: PUB Frequency: 5 Percentage: 0.05 %  Origin: SAV Frequency: 21 Percentage: 0.21 %  Origin: SBA Frequency: 7 Percentage: 0.07 %  Origin: TUL Frequency: 58 Percentage: 0.58 %  Origin: ORH Frequency: 5 Percentage: 0.05 %  Origin: LEX Frequency: 27 Percentage: 0.27 %  Origin: ORD Frequency: 493 Percentage: 4.93 %  Origin: SBN Frequency: 6 Percentage: 0.06 %  Origin: MKE Frequency: 31 Percentage: 0.31 %  Origin: GNV Frequency: 5 Percentage: 0.05 %  Origin: MSY Frequency: 74 Percentage: 0.74 %  Origin: AZO Frequency: 6 Percentage: 0.06 %  Origin: MSP Frequency: 176 Percentage: 1.76 %  Origin: CAK Frequency: 10 Percentage: 0.1 %  Origin: MSN Frequency: 10 Percentage: 0.1 %  Origin: TOL Frequency: 7 Percentage: 0.07 %  Origin: CVG Frequency: 124 Percentage: 1.24 %  Origin: ERI Frequency: 7 Percentage: 0.07 %  Origin: CAE Frequency: 34 Percentage: 0.34 %  Origin: MRY Frequency: 5 Percentage: 0.05 %  Origin: BIL Frequency: 18 Percentage: 0.18 %  Origin: FAY Frequency: 7 Percentage: 0.07 %  Origin: JAX Frequency: 41 Percentage: 0.41 %  Origin: DAB Frequency: 15 Percentage: 0.15 %  Origin: GFK Frequency: 3 Percentage: 0.03 %  Origin: DAL Frequency: 68 Percentage: 0.68 %  Origin: ELM Frequency: 9 Percentage: 0.09 %  Origin: OTZ Frequency: 3 Percentage: 0.03 %  Origin: FAT Frequency: 11 Percentage: 0.11 %  Origin: FAI Frequency: 6 Percentage: 0.06 %  Origin: FLG Frequency: 6 Percentage: 0.06 %  Origin: EAU Frequency: 1 Percentage: 0.01 %  Origin: DAY Frequency: 76 Percentage: 0.76 %  Origin: ONT Frequency: 69 Percentage: 0.69 %  Origin: SGF Frequency: 6 Percentage: 0.06 %  Origin: MLB Frequency: 12 Percentage: 0.12 %  Origin: MCI Frequency: 108 Percentage: 1.08 %  Origin: LGB Frequency: 15 Percentage: 0.15 %  Origin: PNS Frequency: 14 Percentage: 0.14 %  Origin: LGA Frequency: 208 Percentage: 2.08 %  Origin: PFN Frequency: 3 Percentage: 0.03 %  Origin: FAR Frequency: 9 Percentage: 0.09 %  Origin: SDF Frequency: 53 Percentage: 0.53 %  Origin: RST Frequency: 6 Percentage: 0.06 %  Origin: OGG Frequency: 10 Percentage: 0.1 % |

**Destination:**

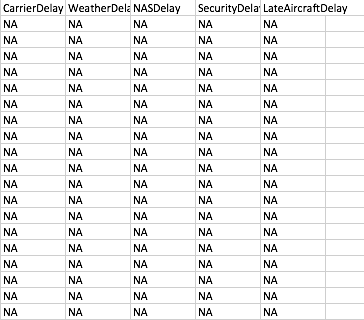
|  |  |
| --- | --- |
| Destination: JFK Frequency: 85 Percentage 0.85 %  Destination: GSP Frequency: 18 Percentage 0.18 %  Destination: FNT Frequency: 3 Percentage 0.03 %  Destination: SIT Frequency: 2 Percentage 0.02 %  Destination: MIA Frequency: 122 Percentage 1.22 %  Destination: BOS Frequency: 203 Percentage 2.03 %  Destination: OAK Frequency: 68 Percentage 0.68 %  Destination: BGM Frequency: 5 Percentage 0.05 %  Destination: CCR Frequency: 3 Percentage 0.03 %  Destination: LIT Frequency: 34 Percentage 0.34 %  Destination: BOI Frequency: 12 Percentage 0.12 %  Destination: BGR Frequency: 6 Percentage 0.06 %  Destination: YUM Frequency: 5 Percentage 0.05 %  Destination: DRO Frequency: 3 Percentage 0.03 %  Destination: CPR Frequency: 6 Percentage 0.06 %  Destination: AGS Frequency: 5 Percentage 0.05 %  Destination: GSO Frequency: 49 Percentage 0.49 %  Destination: SAN Frequency: 109 Percentage 1.09 %  Destination: PIA Frequency: 9 Percentage 0.09 %  Destination: MYR Frequency: 4 Percentage 0.04 %  Destination: PIE Frequency: 1 Percentage 0.01 %  Destination: DCA Frequency: 154 Percentage 1.54 %  Destination: LBB Frequency: 12 Percentage 0.12 %  Destination: HSV Frequency: 12 Percentage 0.12 %  Destination: RAP Frequency: 6 Percentage 0.06 %  Destination: BWI Frequency: 121 Percentage 1.21 %  Destination: PIT Frequency: 231 Percentage 2.31 %  Destination: SAV Frequency: 22 Percentage 0.22 %  Destination: SAT Frequency: 43 Percentage 0.43 %  Destination: CHA Frequency: 10 Percentage 0.1 %  Destination: ILM Frequency: 6 Percentage 0.06 %  Destination: IAH Frequency: 155 Percentage 1.55 %  Destination: JAX Frequency: 32 Percentage 0.32 %  Destination: IAD Frequency: 114 Percentage 1.14 %  Destination: OAJ Frequency: 2 Percentage 0.02 %  Destination: JAN Frequency: 9 Percentage 0.09 %  Destination: HRL Frequency: 14 Percentage 0.14 %  Destination: CHS Frequency: 32 Percentage 0.32 %  Destination: EYW Frequency: 3 Percentage 0.03 %  Destination: WRG Frequency: 4 Percentage 0.04 %  Destination: BNA Frequency: 85 Percentage 0.85 %  Destination: SPN Frequency: 2 Percentage 0.02 %  Destination: JAC Frequency: 1 Percentage 0.01 %  Destination: LAN Frequency: 5 Percentage 0.05 %  Destination: OME Frequency: 1 Percentage 0.01 %  Destination: JNU Frequency: 9 Percentage 0.09 %  Destination: SNA Frequency: 50 Percentage 0.5 %  Destination: PHL Frequency: 152 Percentage 1.52 %  Destination: PHF Frequency: 2 Percentage 0.02 %  Destination: BIS Frequency: 13 Percentage 0.13 %  Destination: SYR Frequency: 59 Percentage 0.59 %  Destination: PHX Frequency: 236 Percentage 2.36 %  Destination: LAX Frequency: 356 Percentage 3.56 %  Destination: MAF Frequency: 23 Percentage 0.23 %  Destination: MBS Frequency: 5 Percentage 0.05 %  Destination: APF Frequency: 2 Percentage 0.02 %  Destination: LAS Frequency: 130 Percentage 1.3 %  Destination: CRP Frequency: 4 Percentage 0.04 %  Destination: CRW Frequency: 9 Percentage 0.09 %  Destination: CMI Frequency: 3 Percentage 0.03 %  Destination: CMH Frequency: 52 Percentage 0.52 %  Destination: GJT Frequency: 3 Percentage 0.03 %  Destination: GUM Frequency: 9 Percentage 0.09 %  Destination: FLL Frequency: 78 Percentage 0.78 %  Destination: DEN Frequency: 362 Percentage 3.62 %  Destination: BZN Frequency: 10 Percentage 0.1 %  Destination: HDN Frequency: 1 Percentage 0.01 %  Destination: DTW Frequency: 221 Percentage 2.21 %  Destination: EVV Frequency: 7 Percentage 0.07 %  Destination: LSE Frequency: 1 Percentage 0.01 %  Destination: MFE Frequency: 3 Percentage 0.03 %  Destination: SFO Frequency: 310 Percentage 3.1 %  Destination: BUR Frequency: 56 Percentage 0.56 %  Destination: ROA Frequency: 11 Percentage 0.11 %  Destination: ROC Frequency: 30 Percentage 0.3 %  Destination: SCC Frequency: 2 Percentage 0.02 %  Destination: MFR Frequency: 8 Percentage 0.08 %  Destination: GEG Frequency: 19 Percentage 0.19 %  Destination: LIH Frequency: 3 Percentage 0.03 %  Destination: ROR Frequency: 1 Percentage 0.01 %  Destination: IDA Frequency: 2 Percentage 0.02 %  Destination: BUF Frequency: 46 Percentage 0.46 %  Destination: ANC Frequency: 20 Percentage 0.2 %  Destination: TRI Frequency: 9 Percentage 0.09 %  Destination: GRR Frequency: 17 Percentage 0.17 %  Destination: SHV Frequency: 15 Percentage 0.15 %  Destination: BDL Frequency: 48 Percentage 0.48 %  Destination: CSG Frequency: 3 Percentage 0.03 %  Destination: DLH Frequency: 5 Percentage 0.05 %  Destination: YKM Frequency: 3 Percentage 0.03 %  Destination: GCN Frequency: 2 Percentage 0.02 %  Destination: LEX Frequency: 26 Percentage 0.26 %  Destination: DSM Frequency: 33 Percentage 0.33 %  Destination: GRB Frequency: 3 Percentage 0.03 %  Destination: EWR Frequency: 234 Percentage 2.34 %  Destination: MHT Frequency: 2 Percentage 0.02 %  Destination: PBI Frequency: 46 Percentage 0.46 %  Destination: MSO Frequency: 1 Percentage 0.01 %  Destination: BTV Frequency: 8 Percentage 0.08 %  Destination: RNO Frequency: 52 Percentage 0.52 %  Destination: BTR Frequency: 14 Percentage 0.14 %  Destination: PSC Frequency: 3 Percentage 0.03 %  Destination: FOE Frequency: 2 Percentage 0.02 %  Destination: AVP Frequency: 4 Percentage 0.04 %  Destination: TVL Frequency: 4 Percentage 0.04 %  Destination: MGM Frequency: 8 Percentage 0.08 %  Destination: MOT Frequency: 2 Percentage 0.02 %  Destination: AVL Frequency: 4 Percentage 0.04 %  Destination: HLN Frequency: 2 Percentage 0.02 %  Destination: OKC Frequency: 46 Percentage 0.46 %  Destination: HTS Frequency: 5 Percentage 0.05 %  Destination: IND Frequency: 60 Percentage 0.6 %  Destination: OGG Frequency: 4 Percentage 0.04 %  Destination: RSW Frequency: 28 Percentage 0.28 % | Destination: MOB Frequency: 16 Percentage 0.16 %  Destination: PSP Frequency: 11 Percentage 0.11 %  Destination: TYS Frequency: 11 Percentage 0.11 %  Destination: FLG Frequency: 3 Percentage 0.03 %  Destination: FWA Frequency: 7 Percentage 0.07 %  Destination: COS Frequency: 22 Percentage 0.22 %  Destination: FSD Frequency: 8 Percentage 0.08 %  Destination: ATL Frequency: 524 Percentage 5.24 %  Destination: ISP Frequency: 14 Percentage 0.14 %  Destination: HNL Frequency: 33 Percentage 0.33 %  Destination: ATW Frequency: 6 Percentage 0.06 %  Destination: ISO Frequency: 8 Percentage 0.08 %  Destination: ACV Frequency: 1 Percentage 0.01 %  Destination: SMF Frequency: 56 Percentage 0.56 %  Destination: SUX Frequency: 12 Percentage 0.12 %  Destination: LFT Frequency: 3 Percentage 0.03 %  Destination: CID Frequency: 4 Percentage 0.04 %  Destination: FCA Frequency: 1 Percentage 0.01 %  Destination: PVD Frequency: 16 Percentage 0.16 %  Destination: SEA Frequency: 123 Percentage 1.23 %  Destination: ICT Frequency: 24 Percentage 0.24 %  Destination: MDW Frequency: 34 Percentage 0.34 %  Destination: MDT Frequency: 12 Percentage 0.12 %  Destination: RDU Frequency: 120 Percentage 1.2 %  Destination: LNK Frequency: 12 Percentage 0.12 %  Destination: GTF Frequency: 7 Percentage 0.07 %  Destination: PDX Frequency: 87 Percentage 0.87 %  Destination: CLE Frequency: 112 Percentage 1.12 %  Destination: DFW Frequency: 402 Percentage 4.02 %  Destination: SJU Frequency: 32 Percentage 0.32 %  Destination: RDD Frequency: 2 Percentage 0.02 %  Destination: AUS Frequency: 67 Percentage 0.67 %  Destination: SRQ Frequency: 23 Percentage 0.23 %  Destination: YAK Frequency: 1 Percentage 0.01 %  Destination: CLT Frequency: 171 Percentage 1.71 %  Destination: SJC Frequency: 78 Percentage 0.78 %  Destination: ELP Frequency: 51 Percentage 0.51 %  Destination: TLH Frequency: 15 Percentage 0.15 %  Destination: ALO Frequency: 8 Percentage 0.08 %  Destination: OMA Frequency: 25 Percentage 0.25 %  Destination: FAR Frequency: 6 Percentage 0.06 %  Destination: BLI Frequency: 2 Percentage 0.02 %  Destination: ITH Frequency: 10 Percentage 0.1 %  Destination: MEM Frequency: 162 Percentage 1.62 %  Destination: LYH Frequency: 1 Percentage 0.01 %  Destination: TUS Frequency: 34 Percentage 0.34 %  Destination: ALB Frequency: 30 Percentage 0.3 %  Destination: PUB Frequency: 1 Percentage 0.01 %  Destination: RIC Frequency: 28 Percentage 0.28 %  Destination: CDV Frequency: 1 Percentage 0.01 %  Destination: SBA Frequency: 5 Percentage 0.05 %  Destination: TUL Frequency: 46 Percentage 0.46 %  Destination: ORH Frequency: 7 Percentage 0.07 %  Destination: ORF Frequency: 62 Percentage 0.62 %  Destination: ORD Frequency: 511 Percentage 5.11 %  Destination: SBN Frequency: 4 Percentage 0.04 %  Destination: MKE Frequency: 36 Percentage 0.36 %  Destination: GNV Frequency: 4 Percentage 0.04 %  Destination: MSY Frequency: 85 Percentage 0.85 %  Destination: AZO Frequency: 7 Percentage 0.07 %  Destination: KOA Frequency: 2 Percentage 0.02 %  Destination: MSP Frequency: 180 Percentage 1.8 %  Destination: CAK Frequency: 14 Percentage 0.14 %  Destination: MSN Frequency: 14 Percentage 0.14 %  Destination: TOL Frequency: 6 Percentage 0.06 %  Destination: CVG Frequency: 109 Percentage 1.09 %  Destination: ERI Frequency: 5 Percentage 0.05 %  Destination: CAE Frequency: 20 Percentage 0.2 %  Destination: MRY Frequency: 9 Percentage 0.09 %  Destination: BIL Frequency: 14 Percentage 0.14 %  Destination: FAY Frequency: 5 Percentage 0.05 %  Destination: CHO Frequency: 3 Percentage 0.03 %  Destination: TPA Frequency: 126 Percentage 1.26 %  Destination: DAB Frequency: 11 Percentage 0.11 %  Destination: GFK Frequency: 2 Percentage 0.02 %  Destination: DAL Frequency: 77 Percentage 0.77 %  Destination: ELM Frequency: 5 Percentage 0.05 %  Destination: OTZ Frequency: 2 Percentage 0.02 %  Destination: FAT Frequency: 7 Percentage 0.07 %  Destination: FAI Frequency: 8 Percentage 0.08 %  Destination: BET Frequency: 1 Percentage 0.01 %  Destination: EAU Frequency: 3 Percentage 0.03 %  Destination: DAY Frequency: 80 Percentage 0.8 %  Destination: ONT Frequency: 54 Percentage 0.54 %  Destination: SGF Frequency: 3 Percentage 0.03 %  Destination: MLB Frequency: 16 Percentage 0.16 %  Destination: ABE Frequency: 14 Percentage 0.14 %  Destination: STL Frequency: 253 Percentage 2.53 %  Destination: MLI Frequency: 12 Percentage 0.12 %  Destination: EUG Frequency: 8 Percentage 0.08 %  Destination: STT Frequency: 13 Percentage 0.13 %  Destination: ABQ Frequency: 50 Percentage 0.5 %  Destination: KTN Frequency: 4 Percentage 0.04 %  Destination: HOU Frequency: 111 Percentage 1.11 %  Destination: HPN Frequency: 2 Percentage 0.02 %  Destination: MLU Frequency: 6 Percentage 0.06 %  Destination: CWA Frequency: 3 Percentage 0.03 %  Destination: BFL Frequency: 5 Percentage 0.05 %  Destination: STX Frequency: 7 Percentage 0.07 %  Destination: SLC Frequency: 128 Percentage 1.28 %  Destination: MCO Frequency: 129 Percentage 1.29 %  Destination: PWM Frequency: 12 Percentage 0.12 %  Destination: BHM Frequency: 45 Percentage 0.45 %  Destination: VPS Frequency: 2 Percentage 0.02 %  Destination: MCI Frequency: 104 Percentage 1.04 %  Destination: LGB Frequency: 7 Percentage 0.07 %  Destination: PNS Frequency: 10 Percentage 0.1 %  Destination: LGA Frequency: 201 Percentage 2.01 %  Destination: PFN Frequency: 2 Percentage 0.02 %  Destination: AMA Frequency: 13 Percentage 0.13 %  Destination: SDF Frequency: 47 Percentage 0.47 %  Destination: PSG Frequency: 1 Percentage 0.01 %  Destination: RST Frequency: 2 Percentage 0.02 % |

**Raw Data:** The data we used came from a CSV file. Below is a small sample of the larger file with which we started. Some of the attributes are clearly not very useful: the year, for example, is going to be 1987 for each entry, since the CSV file is specifically a record from the year 1987. The last five columns are also pretty useless, since they have row upon row of NA, which means they have no value. We chose five columns that had a lot of data so we could come up with good statistics.

**Source web page:** <http://stat-computing.org/dataexpo/2009/the-data.html>







The columns are:

**Name Description**

Year 1987-2008

Month 1-12

DayOfMonth 1-31

DayOfWeek 1 (Monday) - 7 (Sunday)

DepTime actual departure time (local, hhmm)

CRSDepTime scheduled departure time (local, hhmm)

ArrTime actual arrival time (local, hhmm)

CRSArrTime scheduled arrival time (local, hhmm)

Unique Carrier unique carrier code

FlightNum flight number

TailNum plane tail number

ActualElapsedTime in minutes

CRSElapsedTime in minutes

AirTime in minutes

Arry Delay arrival delay, in minutes

DepDelay departure delay, in minutes

Origin origin IATA airport code

Dest destination IATA airport code

Distance in miles

TaxiIn taxi in time, in minutes

TaxiOut taxi out time, in minutes

Cancelled was the flight cancelled?

CancellationCode reason for cancellation (A = carrier, B = weather, C = NAS, D = security)

Diverted 1 = yes, 0 = no

CarrierDelay in minutes

WeatherDelay in minutes

NASDelay in minutes

SecurityDelay in minutes

LateAircraftDelay in minutes